Implicit versus Explicit Mechanisms of Vocabulary Learning and Consolidation

Justyna M. Sobczak[a*], M. Gareth Gaskell[b]

[a]School of Psychological and Social Sciences, York St. John University, York, Lord Mayor's Walk, YO31 7EX, UK
[b]Department of Psychology, University of York, York, Heslington, YO10 5DD, UK

*Corresponding Author: Justyna Sobczak, E-mail: j.needham@yorksj.ac.uk

Previous research has suggested that integration of novel words into lexical competition benefits from a consolidation delay containing a period of sleep (Dumay & Gaskell, 2007). However, a recent study argued that learning novel words via a relatively implicit Hebb repetition task leads to later lexical integration independently of sleep (Szmalec, Page, & Duyck, 2012). It is not clear whether this different time course of lexical integration is a consequence of the learning method chosen, as opposed to other between study differences. Four experiments directly compared the learning of novel words using explicit and implicit methods, namely phoneme monitoring on isolated tokens vs. Hebb repetition of syllable sequences. The impact of the learning was tested at a range of later time-points using two tests of explicit knowledge (recognition and recall) and a test of lexical integration (pause detection on related existing words). Between experiments, we also manipulated exposure frequency, the impact of syllable grouping cues in Hebb repetition and the level of mismatch between novel and real words. The results suggested that learning novel words via Hebb sequence repetition does not confer a benefit on lexical integration prior to or after sleep. We observed an engagement in lexical competition only in the case where a good level of explicit training was followed by a consolidation delay. Recognition and recall performance was generally poorer for Hebb learning. We conclude that Hebb-style implicit learning of words does not allow consolidation processes to be bypassed in lexical integration.

Key words: Hebb repetition task; explicit and implicit learning; lexical integration; lexical competition effect; word learning

Introduction

Language learning is undoubtedly one of the most crucial processes in human development, yet the time-course and mechanisms underlying the establishment of lexical entries are not fully understood. On the one hand there is a well-documented argument in the adult (e.g., Fernandes, Kolinsky, & Ventura, 2009; Kapnoula & McMurray, 2015; Kapnoula, Gupta, Packard, & McMurray, 2015) and developmental literature (e.g., Carey & Bartlett, 1978; Carey, 1978; Spiegel & Halberda, 2011) that phonological forms may be acquired swiftly. On the other hand, there is evidence to suggest that the development of a fully-fledged representation of a novel word may be a more extended process over the course of days or weeks (Bakker, Takashima, Hell, & Janzen, 2015; Bakker, Takashima, van Hell, Janzen, & McQueen, 2014; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003b). To what extent the time-course of novel word learning is modulated by the encoding circumstances is currently under debate.

Successful word learning includes an integration process that allows novel items to gain properties and status similar to established lexical items. Once a novel word has been fully integrated into mental lexicon it should engage in the automatic lexical recognition process whereby it becomes identified in competition with other similarly sounding words (Gaskell & Marslen-Wilson, 2002; Norris, 1994). Research on word learning has indicated that this integration of novel spoken words is typically supported by a consolidation process often associated with sleep (Davis & Gaskell, 2009; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003b; Henderson, Weighall, Brown, & Gaskell, 2012; Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010). For example, Gaskell and Dumay (2003b) and Dumay and Gaskell (2007) investigated the possible role of sleep in lexical integration by teaching their participants fictitious novel spoken words such as *cathedruke* (designed to partially overlap with existing words) and then testing how learning these novel words affected processing of their existing neighbours (e.g., *cathedral*) across different time delays. In an auditory lexical decision or pause detection (Mattys & Clark, 2002) task an increase in response time to the existing word is taken to indicate engagement of the novel word in lexical competition with existing neighbours and therefore some level of lexical integration. Dumay and Gaskell (2007) found no evidence of changes in lexical competition immediately after learning. However, they observed a clear enhanced competition effect after a 12-hr period that included nocturnal sleep but, notably, not after a similar period of wakefulness. This time-course and association between sleep and the lexical integration of novel words can be interpreted within a two-stage account of novel word learning and neurocognitive models of declarative memory formation such as the Complementary Learning Systems framework (CLS, McClelland, McNaughton, & O'Reilly, 1995). The CLS model proposes that new declarative information is initially and temporarily stored using hippocampal mediation (Davis, Di Betta, Macdonald, & Gaskell, 2009) and later becomes hippocampally independent as it is incorporated into existing long-term neocortical memories. Here, sleep provides optimal conditions for such transfer as the cognitive system is offline and not engaged in processing of new information (McClelland et al., 1995). This hippocampal mediation of new memory traces has been supported by the active systems model of sleep-dependent consolidation (Born & Wilhelm, 2012; Diekelmann & Born, 2010; Rasch & Born, 2013). The relationship between lexical integration of novel words and sleep has also been more directly tested, revealing one particular aspect of sleep architecture (sleep spindle activity) that was associated with the emergence of lexical competition (Tamminen et al., 2010).

Based on the above, sleep appears to play a prominent role in consolidation of new lexical knowledge. However, the learning of new vocabulary is not necessarily a homogeneous process. Although, sleep was shown to play an important role in a variety of learning contexts, including relatively implicit word learning from stories (Henderson, Devine, Weighall, & Gaskell, 2015), one may argue that the studies that uncovered a possible role of sleep in novel word integration predominantly relied on explicit learning mechanisms. For example, Gaskell & Dumay (2003b) asked participants to listen for particular phonemes within the novel words, which were presented in isolated form with instructions to memorise the novel words for later test. This is quite an explicit form of tuition, and it is possible that more implicit learning tasks and/or less explicitly segmented speech might recruit different learning mechanisms, which might change the nature of the lexical integration process and reduce the importance of sleep. This possibility has been investigated in a series of studies by Szmalec and colleagues using the Hebb repetition effect. The Hebb paradigm involves gradual learning of serially ordered information via repetition. In an immediate serial recall task, Hebb (1961) presented a specific sequence of digits repeatedly every third trial interspersed with nonrepeating sequences and demonstrated that sequence repetition led to superior recall over time. The Hebb effect is thought to be implicit as it occurs irrespective of awareness (Stadler, 1993). Although this learning effect was originally shown for sequences of digits it has since been successfully used across different modalities and with a range of stimuli such as visuo-spatial (Couture & Tremblay, 2006; Guérard, Saint-Aubin, Boucher, & Tremblay, 2011), pictorial (Page, Cumming, Norris, Hitch, & McNeil, 2006), facial (Horton, Hay, & Smyth, 2008), and tactile sequences (Johnson, Cauchi, & Miles, 2013; Johnson, Shaw, & Miles, 2016).

Szmalec et al. (2009; 2012) explored the Hebb effect in novel word learning and argued that processes underlying sequence learning in the Hebb repetition paradigm are vital in language acquisition (see also Cumming, Page, & Norris, 2003; Page & Norris, 2008) and that the task offers a naturalistic model of learning. Consistent with this argument, impaired Hebb sequence learning has been found in dyslexics (Szmalec, Loncke, Page, & Duyck, 2011, but see Staels & Van den Broeck, 2015). More directly, Szmalec, Duyck, Vandierendonck, Mata, and Page (2009) used a variant of the Hebb procedure to examine the learning of wordlike "chunks" from sequences of nonsense syllables (e.g., *zi-lo-ka-ho-fi-se-be-ru-mo*). The sequences used three trisyllable groupings that were presented in different orders across repetitions. The consistent grouping allowed the trisyllables to become familiar units (e.g., *ziloka, hofise, berumo).* In order to assess this familiarity, they used them in a lexical decision task soon after training. The results showed that the three-syllable groupings extracted from the Hebb sequences were somewhat harder to reject as pseudowords than filler trisyllables suggesting a more wordlike representation. The authors argued that the Hebb repetition procedure reflects the implicit way children learn to segment and sequence words from phonological regularities in their environment (but see also Mosse & Jarrold, 2008). Indeed, this form of implicit learning of linguistic regularities from environment has been previously successfully established by statistical language learning studies (Saffran, 2002, 2003) and suggests that the Hebb effect variant, as a form of a statistical learning, may utilise the same mechanism.

Building on these findings, Szmalec, Page, and Duyck (2012) applied similar experimental procedures to investigate the time course of novel word integration. The researchers presented their participants with visual sequences of 9 consonant-vowel (CV) syllables for immediate serial recall (i.e. sa-fa-ra-sa-la-mo-fi-na-lo). The Hebb sequences were repeated every third trial and again the grouping of the sequences facilitated the extraction of trisyllabic pseudowords (i.e. safara, salamo, finalo). Based on the logic of Dumay and Gaskell (2007), the authors then used pause detection to test whether the novel sequences would show engagement in lexical competition with their existing Dutch counterparts (i.e. safari, salami, finale). As in Dumay and Gaskell, groups were trained either in the morning or the evening, and were tested immediately after training and 12 and 24 hours later.

Diverging from Dumay and Gaskell, both groups showed a similar profile of lexical competition induced by the newly learnt trisyllables. Specifically, lexical competition was not found immediately, but emerged after a 12-hour delay in both groups regardless of whether they slept in the intervening period. This pattern of results suggested that although some time delay is necessary to integrate the new items into lexicon, the time lapse itself is sufficient and there is no need for overnight consolidation. The researchers concluded that the exposure to reoccurring Hebb sequences leads to a formation of lexical representations independently of sleep, in contrast with more explicit learning.

The Szmalec et al. (2012) result in comparison with Dumay & Gaskell (2007) strongly suggests that Hebb repetition and more explicit learning utilize distinct memory systems (cf. Foerde, Knowlton, & Poldrack, 2006). Interestingly, the Hebb repetition effect was shown to be unimpaired in hippocampally amnesic patients (Baddeley & Warrington, 1970; Gagnon, Foster, Turcotte, & Jongenelis, 2004), strengthening the case that Hebb repetition does not rely on the hippocampal complex for learning and so allows swifter (although not immediate) consolidation. At the same time, this sparing of Hebb repetition learning in hippocampal amnesics somewhat weakens the case for it representing the main mechanism for word learning, given that amnesic patients tend to manifest major deficits in novel word learning (Bayley, Reilly, Curran, & Squire, 2008).

A second learning paradigm that may recruit separate neuroanatomical substrates in comparison with explicit encoding is fast mapping (Sharon, Moscovitch, & Gilboa, 2011). Fast mapping was coined as a term to describe how children use mutual exclusivity to identify new word meanings (Carey & Bartlett, 1978), often maintaining this knowledge in memory for several days after very few exposures (Swingley, 2010, but see also Horst & Samuelson, 2008). In a typical fast mapping trial, a novel object is presented alongside an object for which the name is known. If a new word is then heard, the correct association between word and object can then be made simply by ruling out the already known item. Coutanche and Thompson-Schill (2014) examined how fast mapping affects the time-course of novel word integration in comparison with explicit encoding using a semantic decision task (Bowers, Davis, & Hanley, 2005). In the fast mapping condition participants were presented with images of unfamiliar animals together with the well-known ones and asked a question that referred to the new animal by name (e.g., "are the antennae of the torato pointing up?"). In the explicit condition, participants were presented with unfamiliar animals and their names and were asked to memorise the novel names (e.g., "remember the torato"). The semantic decision task showed that fast mapping but not explicit encoding led to slower responses to related existing words (e.g., tomato) 10 minutes later, suggesting that fast mapping supported swift lexical integration (Bowers et al., 2005). Moreover, a second experiment suggested that it was the presentation of the already known item during learning that allowed for the rapid integration effect. This indicates that the presence, or accessibility, of previous knowledge may facilitate and speed up learning of novel information. Additionally, these findings provide further evidence for different mechanisms underlying fast mapping and explicit learning and are in agreement with studies on amnesic patients who, despite hippocampal damage, showed rapid learning of information through fast mapping but not the standard memory tasks (Sharon et al., 2011 although cf. Greve, Cooper, & Henson, 2014; Cooper, Greve, & Henson, 2018).

Although the Hebb repetition task resulted in a substantially different time-course of lexical integration in comparison to explicit tasks, it is worth noting that the picture drawn from standard word learning studies themselves is not entirely straightforward. The progress of engagement in lexical competition for novel words is partly dependent on training properties. Although a large body of evidence supports the argument that newly learnt items engage in lexical competition after sleep, in some cases this effect has been found sooner. For instance, Gaskell and Dumay (2003a) found immediate lexical competition when manipulating the frequency of the items to be learnt. Low frequency items, presented 12 times during the encoding phase, showed no evidence of lexical competition effect when tested on the same day of training or even when re-tested a week later. Conversely, the high frequency items, presented 60 times in training, appeared to engage in lexical competition immediately. Correspondingly, immediate lexical competition was also shown in an artificial language learning paradigm for which training involved extensive exposure to novel items in a continuous stream (Fernandes et al., 2009). These results suggest that substantial exposure to novel items can effectively alter the time course of lexical integration, perhaps due to increased automaticity in the novel word recognition (Geukes, Gaskell, & Zwitserlood, 2015; Tham, Lindsay, & Gaskell, 2015) .

Another factor that appears to influence the time course of novel words integration is their co-presentation with existing words. For example, Lindsay and Gaskell (2009) tested whether exposure to novel words spaced throughout a day would accelerate their integration into the lexicon. The authors found that the competition effects indeed emerged before sleep, but only when the exposure to novel items was interleaved with test phases where phonologically similar existing words were presented. This suggests that the time-course of novel word integration can be changed by spaced interleaving with their existing phonological neighbours during learning (Lindsay & Gaskell, 2009). Similarly, Kapnoula et al. (2015) found an immediate lexical competition effect in the co-activation of novel and familiar words using a visual word paradigm (cf. Weighall, Henderson, Barr, Cairney, & Gaskell, 2016). Therefore, whilst offline consolidation plays a crucial, and perhaps optimising, role in improving automaticity with which novel words are accessed, the process of lexical integration itself seems to follow a more graded curve, often dependent on different factors such as a learning condition (cf. McMurray, Kapnoula, & Gaskell, 2016).

In sum, whilst offline consolidation clearly plays an important role, the process and time-course of lexical integration appear to depend on a range of different factors such as learning and testing conditions. The extent to which different profiles of learning and consolidation are available is a crucial issue to address, so that we understand the mechanisms that support vocabulary acquisition in a natural linguistic environment. However, clear evaluation of the different learning mechanisms is only possible if other potentially confounding factors can be eliminated. Some of the apparent differences between different types of word learning may instead be a consequence of different training properties such as the level of overlap between new and known items, be it semantic or phonological. In the current study, we examined the consequences of novel word learning via Hebb repetition and a more explicit phoneme monitoring task whilst at the same time controlling, as far as possible, for potential confounding factors. We used the time-course of engagement in lexical competition as a measure of lexical integration, alongside other declarative memory tests. If differences in the time course of lexical engagement remain when other factors are controlled, then we can be more confident that tasks exploit different learning mechanisms.

Previous Hebb repetition studies of word learning have differed from more explicit novel word training in potentially important ways such as the number of novel words and the number of presentations. In Szmalec et al. (2012) participants were exposed to 6 novel words twelve exposures each during training. The studies based on the phoneme monitoring task used more words and a higher exposure rate (typically thirty exposures or more; Bakker et al., 2014; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003b; Henderson et al., 2012; Tamminen et al., 2010), with fewer exposures sometimes proving to be insufficient for generating lexical competition effects (Gaskell & Dumay, 2003b). On the basis that a low level of exposure sufficed for Hebb repetition to show interesting effects on lexical competition, we decided to retain this low exposure level for both tasks in Experiments 1 and 2. Given previous studies, this should offer a sufficient level of encoding to induce lexical competition after a delay in the Hebb repetition condition even if this is not necessarily the case in the more explicit condition.

A second important way in which previous studies have differed is the relationship between the fictitious novel words and existing words. In Szmalec et al. (2012) novel words overlapped very closely with their Dutch base words, diverging only in the final vowel (e.g., *bikina* versus *bikini*). In contrast, the studies using more explicit learning methods have tended to use either more substantial final deviations (e.g., the final vowel and consonant, as in *cathedruke–cathedral*) or using embeddings (e.g., *lirmucktoze* embedding *muck*). In principle, this should not matter; after all, real word competitors can differ by as little as a single final vowel (e.g., *window–windy*). That said, having such a small deviation could alter the trajectory of learning or the nature of any lexical competition. It has been shown across several languages, including Dutch and English, that vowel changes in words are more easily relatable to the base words than changes in consonants (Cutler, Sebastián-Gallés, Soler-Vilageliu, & van Ooijen, 2000). This fits with the idea that there may be more leniency in the word recognition system for deviations in vowels than consonants (van Ooijen, 1996). It has also been argued that vowels and consonants have different contributions in early word learning (Nazzi, Gopnik, & Karmiloff-Smith, 2005) and that both play different roles in speech processing and language acquisition, with consonants being more important than vowels at the lexical level (Nespor, Peña, & Mehler, 2003). A single vowel deviation between novel and known items may therefore lead to the novel word being treated as a variant of the existing word (Bürki & Gaskell, 2012) which could change the nature of the learning experience. Therefore, the novel items and English base words used in the present study differed on their final CV syllable (e.g., *bikiso–bikini*), in a similar way to the explicit learning studies. By changing the full final syllable, we put to test whether the Hebb repetition learning extends to these more varied competitors.

A final modification of the Hebb repetition task used in the current study concerned stimulus presentation. In contrast to Szmalec et al. (2012), who presented their stimuli visually, we used auditory stimuli. The reasons for this were twofold: firstly, this helped to avoid any potential cross-modal conflict in the interpretation of consolidation effects (cf. Bakker et al., 2014). Secondly, as the current study used the English language, which has a more complex relationship between spelling and the sound compared with the Dutch language used by Szmalec et al. (2012), abandoning visual presentation allowed us to avoid spelling-pronunciation ambiguity.

We hypothesised that participants who learned novel pseudowords via the Hebb repetition task would show lexical integration of novel items after a delay but without needing sleep, similar to the results in Szmalec et al. (2012). It was less clear whether the exposure level would be sufficient for participants who learned novel items via the phoneme monitoring task to show lexical integration of new items (Gaskell & Dumay, 2003a, 2003b), but if there was an effect we expected that this would be strongest after sleep (Dumay & Gaskell, 2007). With regards to the explicit declarative memory tests, our prediction was that learning via a more explicit phoneme monitoring task would result in a more robust declarative memory for novel words (recognition and cued recall tests), in comparison to a more implicit Hebb repetition task, due to the recruitment of attention and conscious control, as a function of training condition (Batterink, Reber, & Paller, 2015).

Experiment 1

Method

Experiment 1 hypotheses, design, procedures and planned analyses were subject to pre-registration at the Open Science Framework (https://osf.io/6p9my/), with some minor alterations noted below. Furthermore, a planned vigilance task was initially included, but was later removed from the experiment due to repeated software failure.

The overall procedure for Experiment 1 is illustrated in Figure 1. Participants attended Session 1 in the morning when they completed either the phoneme monitoring or the Hebb repetition task as a way of familiarising themselves with the novel sequences (e.g. a novel word *bikiso* pronounced as *bih-kee-soo*). The effect of exposure on the lexical competition process for neighbouring existing words (e.g., *bikini*) was then tested using a pause detection task immediately after training. Participants completed another pause detection tasks in the evening, after a 12-hour delay. The third lexical integration test was completed next morning, 24 hours after encoding, following a night of sleep. This experimental design was motivated by the fact that the main interest here was to assess the emergence of lexical competition in the Hebb repetition condition after a delay without sleep. Apart from the lexical integration task there were also explicit tests of novel sequence knowledge: cued recall and recognition tasks, which took place only after the 24-hour delay. In the cued recall task participants heard the first CVC of the novel words and were asked to recall the novel sequences they learnt on the previous day. In the recognition task participants were required to pick up the familiar novel words from spoken pairs differing only in their final syllables (e.g., *bikiso* vs. *bikita*).

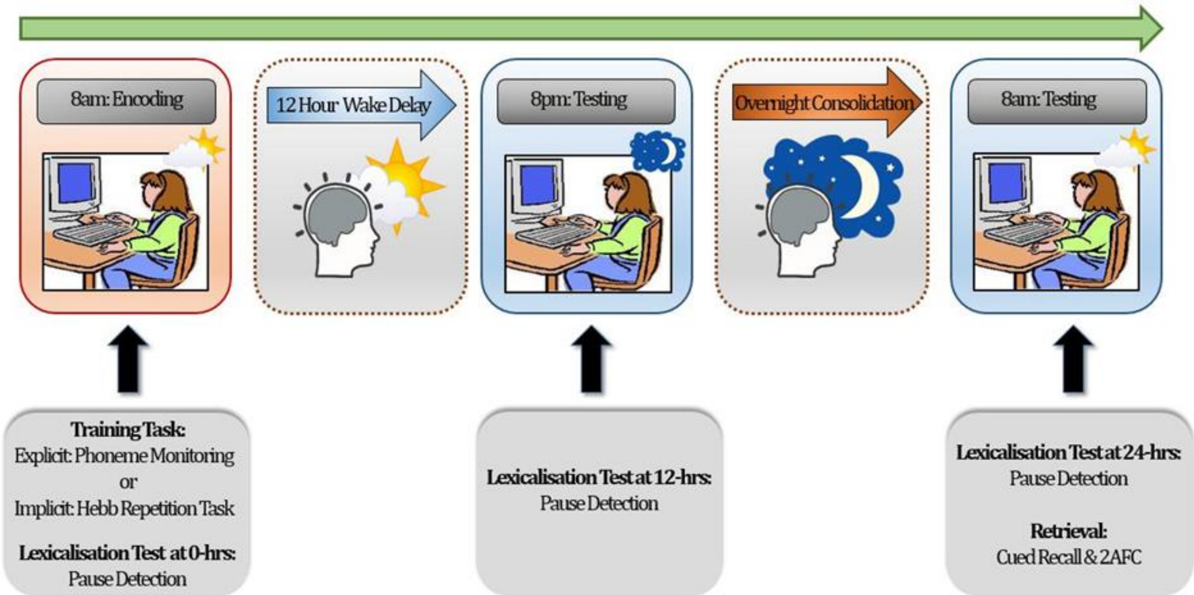


Figure 1. Study procedure for all participants. The encoding phase took place in the morning when participants completed either the Hebb repetition or phoneme monitoring task. The lexical integration test was administered at three time points: immediately after learning (0-hr delay), 12 hours from the learning phase (12-hr delay; during this time participants were instructed to refrain from taking naps) and 24 hours after training (24-hr delay, following nocturnal sleep). The final session also consisted of explicit tests: Cued Recall and 2AFC.

Participants. Forty-eight students (forty-one females), between 18 and 26 years old, (mean age: 19.6 years), participated in this experiment. The preregistration stated 44 participants, but four participants were excluded from analyses at the encoding stage, due to either equipment failure or more than 50% incorrect trials in the training task, and so were replaced. Participants in all experiments reported in this paper were University of York students and participated for course credit or financial reward (£6/hour). All reported English as their first language and had no self-reported diagnoses of hearing problems or developmental language disorder (e.g. dyslexia). All participants were informed about the nature of the tasks and their right to withdraw from the study at any time without penalty. All participants provided written consent before the experiment and were debriefed at the end of it. All experiments received ethical approval from the University of York Psychology Department Ethics Committee.

Materials and design. The novel sequences were designed so as to parallel the materials used in Szmalec et al. (2012) unless there was a clear reason to deviate. In contrast to Szmalec et al., (2012) where participants learnt 6 novel words, in our experiment we doubled this number. This was intended to improve statistical power and generalizability. We therefore created 24 trisyllabic CVCVCV novel pseudowords that overlapped phonologically with existing English base words, with the intention that these could become new cohort competitors to the English words (Gaskell & Marslen-Wilson, 2002). In contrast to Szmalec et al. (2012), the English base words and novel pseudowords differed in their final consonant and vowel to increase the phonological contrast between the two. For example, for the English base word *bikini* we created a novel pseudoword *bikiso* (see Appendix A for a complete list of English base words and novel pseudowords). The 24 base words were all nouns ranging in frequency (SUBTLEX, Brysbaert & New, 2009) between 0.35 and 20.37 occurrences per million (mean: 4.44) and their uniqueness point was always located between the third and fifth phonemic position (Celex; Baayen, Piepenbrock, & van Rijn, 1993). The novel pseudowords retained the stress pattern of their English base words, with primary stress falling on either the first or the second syllable. All materials were recorded in a soundproof booth by a native speaker of British English (MGG). The novel pseudowords were recorded both as continuous trisyllabic forms and as three separate syllables for use in the Hebb repetition task. Care was taken to ensure that the vowels of the separate syllables matched those of the trisyllabic sequence. The sound files were normalised for maximum amplitude and all editing was performed in the Adobe Audition software (Adobe version 3.0).

The test items were then divided into two equal lists which were matched pairwise on the frequency of their base words. During training, participants heard 12 novel items (from one list, counterbalanced across participants). During the lexical integration test participants heard all 24 English base words; half of these had potentially acquired a new competitor (competitor condition) and the other half had not (control condition). This allowed estimation of the speed of recognition for each English base word, with and without influence of the novel competitor.

Participants were allocated randomly to one of two training procedures. In the phoneme monitoring task the novel words were heard as single trisyllabic forms. In the Hebb repetition task the novel words were presented as sequences of syllables and were arranged specifically so that no syllable was repeated within one Hebb sequence of three trisyllable groupings (see Appendix A).

Procedure. The experiment spanned three sessions (see Figure 1). The first and third sessions were administered between 8 and 9 am and the second session between 8 and 9 pm. In the first session participants were exposed to novel sequences in either the phoneme monitoring or the Hebb repetition task. The first session took approximately 1 hour to complete for participants in the Hebb repetition group or 20 minutes for participants in the phoneme monitoring group. Participants returned to the laboratory after a 12-hour break for Session 2 and were instructed to refrain from taking a nap during that time. In the second session participants completed the pause detection task (in a 10-minute session). After another 12-hour break, this time including a normal night's sleep, the third session took place. Participants completed the pause detection task for a third time, followed by two tasks that measured the explicit knowledge of novel pseudowords: cued recall and 2-alternative forced-choice (2AFC). Stimulus presentation over high-quality headphones, timing and data collection were controlled using DMDX (Forster & Forster, 2003), excluding the Hebb repetition task which was presented using E-Prime software. In the *phoneme monitoring* task participants listened to each novel pseudoword and indicated whether a pre-specified phoneme (one of /p/, /n/, /d/, /r/, /m/ and /l/) was present. The target phoneme was the same throughout a block and specified on each trial by displaying the corresponding letter on the screen. The task was preceded by four pseudoword practice trials. Each item occurred 12 times, once per block and twice per target phoneme. The order of the novel pseudowords was randomised within a block. Participants were instructed to respond as quickly as possible by pressing one button if the target was present at any location in the words or press another if it was absent. 250 ms after their response, or after 5,000 ms time-out, the next trial began. As is typical with these experiments, participants were explicitly instructed to try and memorise the novel pseudowords as well as possible in preparation for future tests and to treat them as they were real words of English.

In the *Hebb repetition task* participants listened to ordered sequences of nine syllables. Importantly, care was taken to promote the implicit nature of the task, thus participants were not given any instruction relating to segmentation or chunking of the sequence, or to treat the items as real words. Each participant completed four blocks of 36 sequences each. In each block there was one Hebb sequence (containing three novel pseudoword sequences) presented repeatedly every third trial (12 times in total), and 24 filler sequences. Following the Hebb learning protocol (Couture & Tremblay, 2006; Page et al., 2006; Horton et al., 2008; Guérard et al., 2011, Johnson et al., 2013) all nine syllables were presented consecutively one after another with 500 ms breaks in between. As in Szmalec et al. (2012), but in contrast to the majority of Hebb learning studies, the presentation of the three trisyllable groupings was permuted pseudorandomly. For example, the sequence “mih-mow-lee-row-zuh-no-lih-bee-may” could also be presented as: “row-zuh-no-lih-bee-may-mih-mow-lee”). The order of the syllables in sequences constituting the novel trisyllabic pseudowords was always preserved (e.g., “mih” was always followed by “mow” and then “lee”). It is worth noting that this approach deviates from the original Hebb learning protocol where all elements in the Hebb sequence are presented in the same order on each repetition. Previous research showed that changing the order of the Hebb sequence can affect learning of that sequence. For example, Schwartz and Bryden (1971) indicated that changing the first items in a Hebb sequence has potential to abolish learning of that sequence. Nonetheless, Szmalec et al. (2012) successfully used this altered Hebb protocol in their study and demonstrated a robust Hebb sequence learning.

There were three practice trials at the beginning of the task, after which there was a pause when participants could ask questions. Each trial was followed by an immediate serial recall screen where participants were required to recall verbally the nine syllables in the sequence they were presented and then press the spacebar to move to the next trial. Their responses were recorded and later scored for accuracy. A sheet of paper with nine empty grids was provided to participants to help keep track of the number of syllables they were recalling. They were instructed to say “blank” if they could not recall a particular syllable in a sequence. Overall, participants learned four critical sequences through Hebb repetition across the session, each consisting of three trisyllable groupings that overlap with existing English words (see Figure 2 for a typical trial design). The nonrepeated filler sequences were constructed from different syllables than the Hebb sequences and presented in a random order on each filler trial.

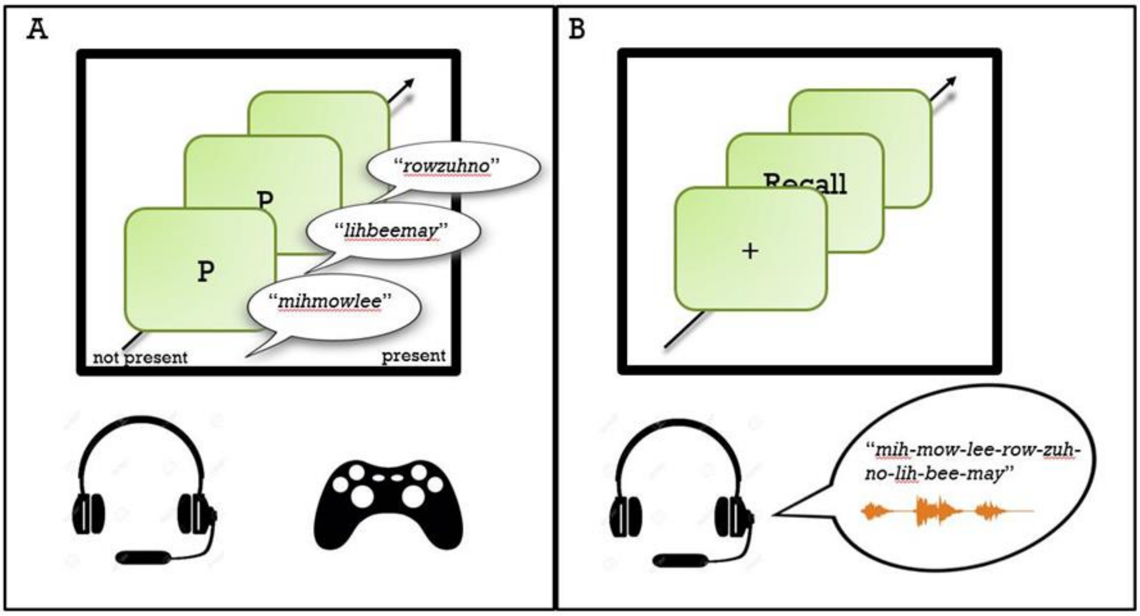


Figure 2. An illustration of three learning trials in the phoneme monitoring (A) and one learning trial in the Hebb Repetition Task (B).

Each of the training sessions was followed by the *pause detection task*, which was intended as a measure of the extent to which the novel sequences had become lexical competitors to the base words and so could influence their recognition. Participants were required to make a speeded decision indicating whether a pause was present in each spoken stimulus by pressing one of two buttons. Stimuli comprised 24 existing words (12 with and 12 without novel competitors) and 56 fillers (40 of CVCVCV structure and 16 of a different structure). Half of the items contained a 200 ms pause inserted directly before the final CV (e.g. *biki_ni*). Four versions of the task were developed and counterbalanced across participants so that each item was equally represented in the four cells of the design (competitor, pause present; competitor, pause absent; control, pause present; control, pause absent). To encourage lexical processing, fillers were all existing words and half of them had a pause inserted at random locations. Response latencies were measured from the alignment point in the waveform that was used to mark pause onset. Participants had six seconds from stimulus onset to respond and each trial was preceded by a cross that appeared on the monitor for 500 ms. The trials were presented as a single block, ordered randomly for each participant. The task started with four practice trials.

In *cued recall* a stem completion test was used. During a typical trial participants heard the first three phonemes (e.g., *bik*-) of the novel pseudowords from the exposure phase and were prompted by a cross on the screen to complete the sequence aloud using one of the new words they had encountered the previous day. Participants in the Hebb repetition condition were asked to recall the syllable sequences that were repeated more frequently than the other in the Hebb repetition task and finish the stem with the matching item. The time between the offset of the cue and the onset of the cross was 500 ms. The cross symbol remained on the screen for 6,000 ms to permit a verbal response before the next trial began. There were 12 randomised trials, each cueing one of the trained pseudowords.

In the final *2AFC* test, participants heard two sequences: a novel pseudoword and its corresponding foil. The foils were constructed in a way that they differed from the novel word, and also its English base word, in their final syllable. For example, the novel word *bikiso* had the foil *bikita.* Participants listened to both sequences before responding with a button press to indicate which sequence had been heard during training. Participants saw an asterisk, displayed on the screen for 500 ms, and then heard the first sequence. After a 500 ms interval the second sequence was played followed immediately by a response instruction. Participants had 5,000 ms to make their response and were instructed to respond as quickly as possible. The order of novel pseudoword/foil pairs was randomised across trials and so was the order of items within each pair. The third session took approximately 20 minutes to complete.

Results

Data from 44 out of 48 participants were entered in analyses as described above, with 22 participants in each training condition.

In the *phoneme monitoring* task, all remaining participants scored at least 83% correct (mean 90%, *SE*= 1%). Of the error responses 6% were misses and 3% were false positive. There was no significant group difference across the experimental lists (p=.752). In the *Hebb repetition task*, as per standard Hebb learning protocol, a CV was scored as correct when recalled in the correct position in the sequence. For each individual participant, regression slopes were calculated for the effect of block on the Hebb sequences and filler sequences. Learning would be reflected in a steeper slope for the Hebb sequences. The gradient values were entered into a one-way repeated measures analysis of variance (ANOVA) with sequence type (filler versus Hebb) as the independent variable. There was a significant main effect of sequence type ($F(1,21)=38.44$, $p=.001$, $\eta p^2=.66$) indicating higher improvement-gradient for Hebb sequences (M=.025, *SE*=.004) relative to fillers (M=.002, *SE*=.001). Therefore, the Hebb effect was obtained, which is a necessary precondition for considering the results of the pause detection task and the explicit tests (see Figure 3). We also assessed the difference between the learning curves of the four blocks. This was not significant, suggesting that the four Hebb sequences were learned at similar level (see Supplementary Materials for more details).

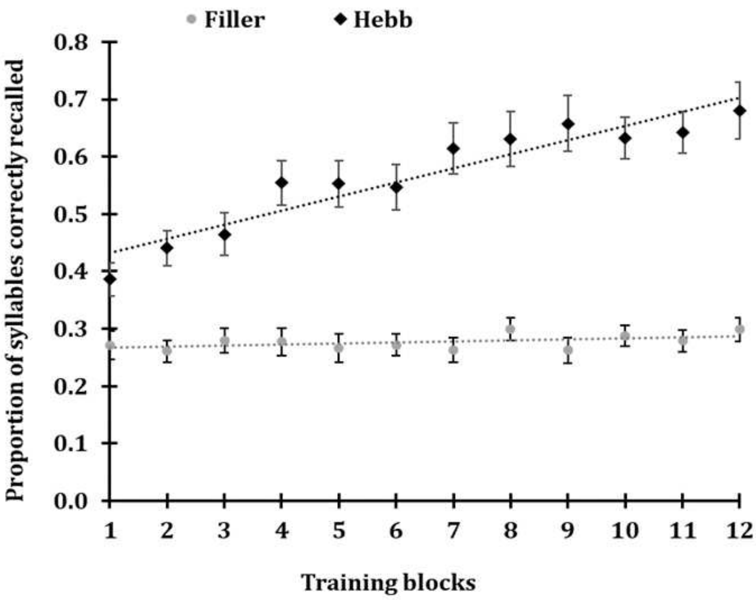


Figure 3. Accuracy (proportion correct) for Hebb and filler sequences in the Hebb repetition task (error bars depict standard error; regression lines illustrate the gradient of improvement in performance).

Pause detection. The data from two participants, one in the Hebb repetition task group and one in the phoneme monitoring group, were excluded from analyses due to more than 33% of incorrect responses. RTs associated with errors, plus all RTs below 150 ms (Tamminen et al., 2010) or above 1,700 ms (Bakker et al., 2014) were removed from the data set (3% for the Hebb repetition task and 4 % for the phoneme monitoring). The RT and error data for experimental items are summarised in Table 1. The reported analyses focused on RTs, as is standard for this type of dependent variable.

Table 1
Mean Pause Detection Latencies (ms) and Error Percentages for Competitor and Control Conditions in Experiments 1 and 2.

		Experiment 1				Experiment 2	
	Training	Hebb Repetition Task		Phoneme Monitoring		Hebb Repetition Task	
	Condition	Competitor	Control	Competitor	Control	Competitor	Control
RT	0-hr	748 (36)	772 (40)	741(31)	734(32)	637 (23)	628 (30)
	12-hr	675 (32)	693 (33)	679(37)	644 (33)	546 (22)	556 (33)
	24-hr	669 (34)	656 (27)	672 (33)	665(36)	520 (21)	527 (27)
% Err	0-hr	1.6 (0.9)	2.4 (1.0)	1.2 (0.6)	1.6 (0.9)	9.0 (1.6)	5.3 (1.3)
	12-hr	0.4 (0.3)	2.8 (0.9)	2.8 (1.1)	2.0 (1.0)	9.0 (1.4)	7.9 (1.4)
	24-hr	1.2 (0.6)	0.8 (0.5)	5.2 (1.9)	1.9 (1.3)	7.2 (1.8)	7.2 (1.7)

Note. Standard error of the mean in parentheses.

RTs for pause present and pause absent trials were averaged across both trial types and RTs were analysed only for correct responses. The latencies were entered into a 2 (training task; phoneme monitoring and Hebb repetition task) × 3 (Session; 0-hr, 12-hr, 24-hr) × 2 (Competitor acquisition: competitor versus control), ANOVAs by participants and items (note that the items analyses were inadvertently left out of the pre-registration document, but are standard in this type of experiment). The analyses revealed that responses became faster over sessions ($F_1(2,80)=20.10$, $p<.001$, $\eta p^2=.334$, $F_2(2,92)=89.86$, $p<.001$, $\eta p^2=.661$) but there was no significant difference in responses in the competitor and control condition (Competitor acquisition, $F_1(1,40)=.16$, $p=.695$, $\eta p^2=.004$, $F_2(1,46)=.039$, $p=.846$, $\eta p^2=.001$). The interactions Session x Training, Session x Competitor acquisition and Session x Competitor acquisition x Training were nonsignificant ($F_1(2,80)=.69$, $p=.503$, $\eta p^2=.017$; $F_2(2,92)=2.37$, $p=.099$, $\eta p^2=.049$; $F1(2,80)=.76$, $p=.471$, $\eta p^2=.019$, $F_2(2,92)=.81$, $p=.448$, $\eta p^2=.017$ and $F_1(2,80)=1.50$, $p=.232$, $\eta p^2=.036$, F_2 $(2,92)=0.59$, $p=.556$, $\eta p^2=.013$ respectively). The Competitor acquisition x Training interaction was also nonsignificant ($F_1(1,40)=3.99$, $p=.053$, $\eta p^2=.091$, $F_2(1,46)=2.45$, $p=.124$, $\eta p^2=.051$), albeit with a slight trend towards overall stronger competition effects for phoneme monitoring than for Hebb training. The between participants factor Training was also nonsignificant ($F(1,40)=.11$, $p=.746$, $\eta p^2=.003$). The magnitude of the differences in the RTs to test and control base words are shown in Figure 4.

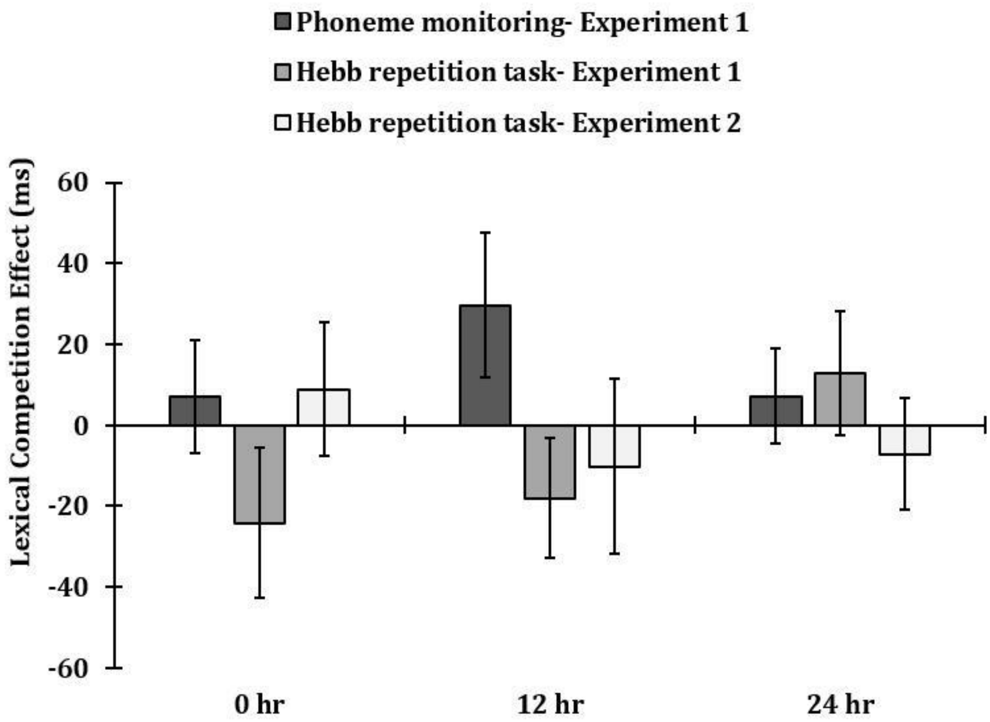


Figure 4. Lexical competition effect (competitor RT- control RT) across three sessions for phoneme monitoring (phoneme monitoring) and Hebb repetition (Hebb repetition task) groups in Experiment 1 and Experiment 2. Error bars represent standard error of the means and are not adjusted to facilitate within-participants' comparisons, given the mixed design (Cousineau & Brien, 2014).

In sum, the Hebb repetition and the phoneme monitoring groups did not show evidence of the lexical competition after delay, regardless of whether the delay contained sleep or not.

Cued recall. In the cued recall task, responses were scored as accurate if the first and middle syllables together with a final consonant were correct (for example, for the novel word *bikiso* the responses: *bikiso*, *bikisoo, bikisa* were all scored as correct but not *bikiro*). This scoring system was motivated by two factors. Firstly, consonants arguably play a more important role in the acquisition and representation of words (Nazzi et al., 2005; Nespor et al., 2003). Secondly, participants' responses during the Hebb repetition indicated that there was some inconsistency in how participants encoded the novel words in the first place, but as long as the sequence prior to the final vowel is encoded then a new competitor to the existing word has been encoded.

Participants' errors mostly involved the final syllable being replaced by the final syllable of another novel pseudoword or the final syllable of the base word. Performance in the cued recall task was relatively poor compared to other published studies (for comparison: above 40% Weighall et al., 2016; above 50% after 24 hrs in Henderson et al., 2013), with participants recalling 17% of the words heard in the training in the Hebb repetition group and 21% in the phoneme monitoring group (see Figure 5). The performance difference between the two groups was not significant ($t_1(40) = .59$, $p = .554$; $t_2(46) = .81$, $p = .421$).

2AFC. Mean accuracy and RT scores for the *2AFC* are presented in Figure 5. Participants recognised the novel pseudowords at a level significantly above chance in both groups (Hebb repetition task: $t_1(20) = 22.47$, $p < .001$, $t_2(23) = 3.423$, $p = .002$; phoneme monitoring: $t_1(20)$ =64.90, p<.001, $t_2(23)$ =23.78, p<.001), with the phoneme monitoring group significantly more accurate than the Hebb repetition group ($t_1(29.43)$=8.96, p<.001, $t_2(46)$=5.96, p<.001). Comparison of the RTs showed that the phoneme monitoring group was significantly faster than the Hebb repetition group in recognising the novel phonological forms ($t_1(40)$=3.56, p=.001, $t_2(46)$=-4.21, p<.001).

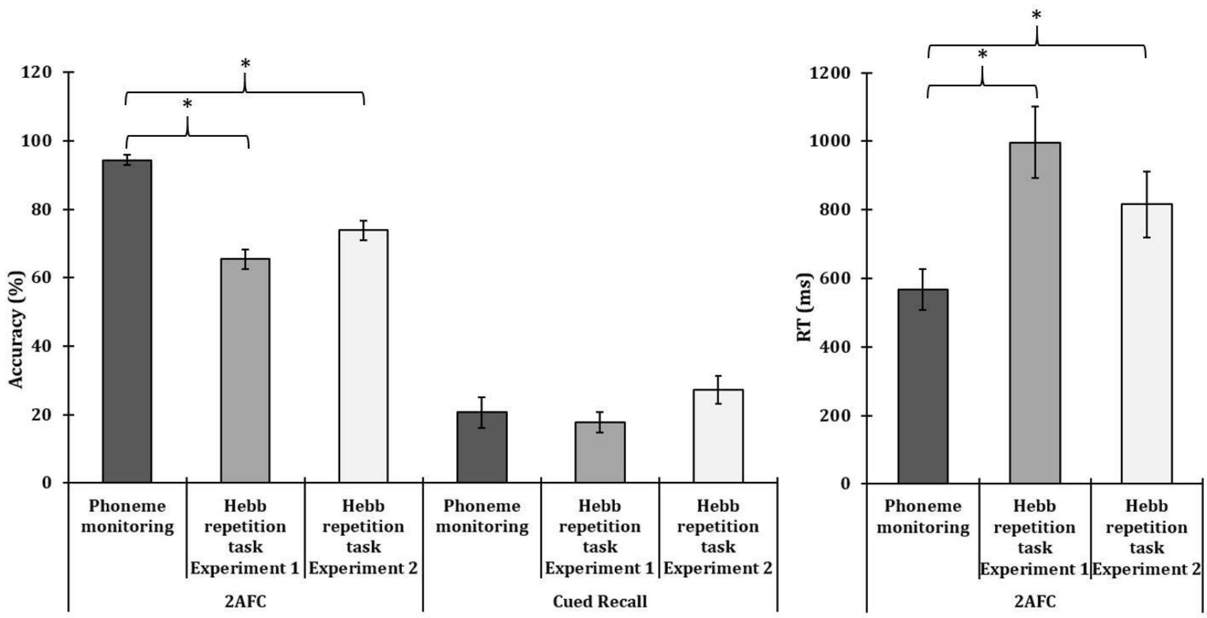


Figure 5. Mean percent correct on explicit tests for the Hebb repetition task and phoneme monitoring groups and mean RTs for both experimental groups in 2AFC task. Error bars represent standard error.

In sum, although both groups recalled the novel pseudowords at a low level (roughly 2 out of 12 words) the phoneme monitoring group showed superior direct recognition of the novel items in comparison to the Hebb repetition group.

Discussion

Experiment 1 compared the changes in dynamics of lexical competition between newly learned phonological forms and their English counterparts after two training tasks, the phoneme monitoring and the Hebb repetition task. We tested the integration of novel words at three time delays: immediately after training, after 12 hours wake period and after 24-hour period, allowing for an overnight sleep. The primary aim was to examine whether the Hebb repetition task, compared with more explicit learning, provides an opportunity for novel words to be better integrated with long-term lexical knowledge prior to sleep, as argued by Szmalec et al. (2012). The results did not support this hypothesis: there was no evidence of an engagement in lexical competition after learning via Hebb repetition. In fact, we did not observe lexical competition effects in either of the groups and regardless of whether or not the time delay included nocturnal sleep. Although the lack of lexical competition effects at any time point in the Hebb condition was a surprise, the lack of an effect for the more explicit learning condition was less so. We chose to match the level of exposure in both conditions to the relatively low level from Szmalec et al. (2012), given that this was sufficient in their Hebb paradigm. The prior evidence relating to this exposure level in explicit learning is more equivocal. For example, Gaskell and Dumay (2003b) did not find a lexical competition effect 24 hours post training when their participants were exposed to novel items 12 times, despite good recognition of the novel forms (as measured by 2AFC task). A second training session with 12 more exposures also did not lead to competition effects after a further 24 hours. Gaskell and Dumay found that the lexical competition effect only emerged after a third session, meaning a total exposure rate of 36 presentations. Later studies showed that an exposure rate of 36 allowed for the lexical competition to emerge after a time-course of 12 and 24 hours, provided that the delay contained sleep (Dumay & Gaskell, 2007; Dumay, Gaskell, & Feng, 2004). In other circumstances, however a lower exposure level seemed to be sufficient. Davis et al. (2009) found a somewhat weak lexical competition effect precisely after 12 presentations, although with a different lexical integration test (i.e. lexical decision) from the current one. It seems likely, given the current results that in an explicit learning task a relatively high level of exposure is needed to guarantee robust evidence of an impact of the novel words on the recognition of their existing neighbours, but that individual differences might contribute to the observation of an effect after weaker exposure in some cases.

As shown by the cued recall task, explicit knowledge of the novel items did not differ between two groups, which is in disagreement with our prediction based on prior studies. It was expected that the novel items encoded via the phoneme monitoring task would be better recalled than those encoded via Hebb repetition. After all, the phoneme monitoring training presented the novel items in isolation with a direct instruction to retain the forms, whereas Hebb repetition used long equally spaced sequences of isolated syllables and no explicit instruction to group the syllables or retain them in the longer term. However, both groups recalled approximately 20% of novel words. Previous studies that used explicit learning tasks typically showed above 40% accuracy in recall tasks (see Henderson et al., 2013; Weighall et al., 2016 for comparison). This indicates relatively poor knowledge of novel items in both our experimental groups, most likely attributable to the lower level of exposure in this experiment. Nonetheless, as we predicted, the easier 2AFC recognition test revealed that the group that learned novel pseudowords via the phoneme monitoring performed significantly better than the group that learned via the Hebb repetition task. This indicates that learning via the Hebb paradigm may lead to less explicit awareness of the repeated sequences.

Given that we did not find the expected impact of Hebb repetition learning on lexical competition, an obvious follow would be to increase the exposure level in training to a level at which we can be confident that explicit training will lead to lexical competition (e.g. Gaskell & Dumay, 2003b). The key question would then be whether Hebb repetition also shows lexical competition. However, one other possible explanation for the lack of a lexical competition effect was worth consideration. As in the standard Hebb learning protocol, Experiment 1 presented trials containing the three trisyllable sequences with no temporal cues to grouping. The desired grouping into trisyllables could only be determined from the transitional probabilities of syllable pairs (Saffran, Aslin, & Newport, 1996; Pelucchi, Hay, & Saffran, 2009; Saffran, 2002, 2003; Saffran, Senghas, & Trueswell, 2000) across Hebb blocks, due to the reordering of these fixed trisyllables in every Hebb block. However, the Hebb trials used by Szmalec et al. (2012) included a more overt cue to aid segmentation: 2,000 ms gaps between the three-syllable groupings. This methodological detail was not reported in Szmalec et al. (2012) but was clarified to us later by one of the authors. These quite long gaps could have both positive and negative aspects. In terms of segmentation, these grouping cues most likely helped to chunk the 9-syllable sequences into the appropriate word-like units. From this point of view, the cues would strengthen the ability of the implicit mechanisms underlying Hebb learning to acquire the appropriate phonemic sequences. At the same, the cues may increase awareness of the groupings as separable strings, perhaps reducing reliance on implicit learning mechanisms and increasing reliance on explicit mechanisms. Therefore, in Experiment 2 we examined whether the inclusion of these temporal chunking cues alters the pattern of lexical engagement in the Hebb training condition. With regards to tests measuring explicit knowledge we predicted that the clearer chunking cues would enhance the declarative memory of the novel phonological forms.

Experiment 2

Method

In Experiment 2 we addressed the influence of the inclusion of temporal grouping cues in the Hebb repetition task on the lexical integration of novel items. Because the temporal grouping variable is only relevant to the Hebb effect style of learning, the phoneme monitoring condition was dropped for Experiment 2.

Participants. Twenty-two participants (15 females), aged between 18 and 25 (mean age 20.2 years), who hadn't taken part in Experiment 1, were trained on novel items using a new version of the Hebb repetition task. The criteria for participation were the same as in the previous experiments.

Material, design and procedure. The critical stimuli were the novel items used in Experiment 1. This time however, 2,000 ms silent gaps were inserted between the three trisyllable groupings constituting the Hebb and the Filler sequences (e.g., "*mih-moh-lee (...) roh-sah-noh (...) lih-bee-may*"). The inclusion of the gaps was likely to make the grouping in the Hebb repetition task more transparent than in Experiment 1, and so unlike Experiment 1, here we added an awareness questionnaire. Upon completion of the experiment, participants filled out a post-experimental questionnaire to assess their awareness of list repetition in the Hebb task and the objective of the experiment. In the questionnaire, participants were asked to state what they thought was the purpose of the experiment and to report any patterns they may have noticed to determine whether they have any explicit knowledge of the re-occurring Hebb sequences. If, in their responses, participants stated novel word learning, word sequence learning, learning via repetition as a purpose of the experiment and/or said that they notice the repeating sequences in the Hebb task they were classed as aware.

The experimental design, procedure and the experimental tasks were otherwise identical to the Hebb condition of Experiment 1.

Results

In the Hebb repetition task, the recall accuracy and regression slopes were calculated according to the previously outlined criteria. The gradient values were entered into a one-way repeated measures ANOVA with sequence type (Filler versus Hebb) as the independent variable. There was a significant main effect of sequence type, $F(1,21)=38.96$, $p<.001$, $\eta p^2=.65$ indicating a higher improvement-gradient for Hebb sequences (M=.025, SE=.004) relative to fillers (M=.004, SE=.001). Therefore, the Hebb effect was again obtained (see Figure 6). Similar to Experiment 1, we assessed the difference between the learning curves of the four blocks which was not significant (please see the supplementary materials for more details).

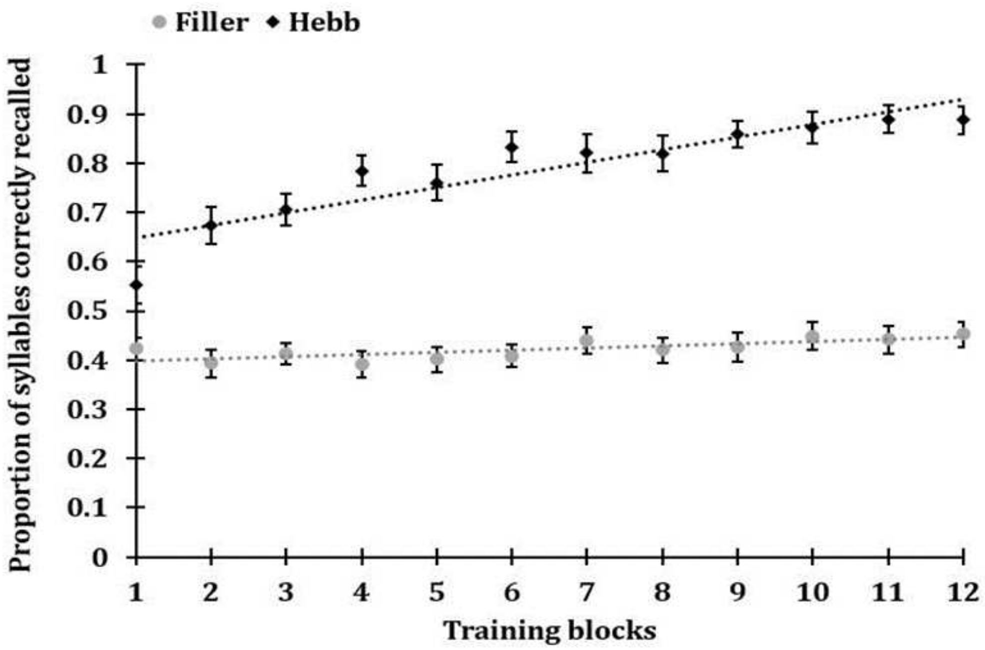


Figure 6. Accuracy (proportion correct) for Hebb and filler sequences in the Hebb repetition task. Values for filler trials represent the average of the two filler sequences presented between each of Hebb sequences (error bars depict standard error; regression lines illustrate the gradient of improvement in performance).

Participant Awareness. Of the twenty-two participants, fourteen (64%) were classified as being aware of syllable repetition and the study's aim on the basis of the post-experimental questionnaire. In their answers, participants either stated that they were aware of the syllable strings constituting novel pseudowords or that the purpose of the Hebb repetition task was to learn novel words. Participants also listed some of the novel words in a syllabic form as examples.

Pause detection. Mean RTs and error data for experimental items are summarised in Table 1. The RT data were analysed using the same methodology and data exclusion criteria (9%) as in Experiment 1using 3 (Session; 0-hr, 12-hr, 24-hr) × 2 (Competitor acquisition: competitor versus control), repeated measures ANOVAs. The analyses revealed that responses became faster over sessions ($F_1(2,42)=16.69$, $p<.001$, $\eta_p^2=.443$, $F_2(2,46)=63.12$, $p<.001$, $\eta_p^2=.733$). There was no significant difference in responses in the competitor and control condition ($F_1(1, 21)=.052$, $p=.822$, $\eta_p^2=002$, $F_2(1,23)=2.76$, $p=.110$, $\eta_p^2=.107$). The interaction Session x Competitor acquisition ($F_1(2,42)=.53$, $p=.593$, $\eta_p^2=.025$, $F_2(2,46)=.04$, $p=.957$, $\eta_p^2=.002$) was nonsignificant (see Figure 4).

Cued Recall and 2AFC. Responses in the cued recall task were scored as in Experiment 1. The inclusion of gaps between the virtual pseudowords appeared to result in more items being recalled (27%) in comparison with the Hebb condition of the previous experiment, although this difference did not reach significance level ($t_1(41)=-1.95$, $p=.058$; $t_2(34.06)=-2.02$, $p=.051$). A cross-experiment comparison of recall accuracy also revealed no difference between the phoneme monitoring group from the Experiment 1 and the Hebb group in Experiment 2 ($t_1(41)=-1.18$, $p=.270$, $t_2(46)=-1.38$, $p=.175$).

In the *2AFC* task participants recognised the novel pseudowords at a level significantly above the chance ($t_1(21)=25.61$, $p<.001$, $t_2(23)=5.20$, $p<.001$). Comparison between Experiment 1 and Experiment 2 in the explicit recognition test showed that despite an increased recognition level in the Hebb group in Experiment 2, the phoneme monitoring group still recognised significantly more items ($t_1(41)=6.32$, $p<.001$, $t_2(46)=4.15$, $p<.001$) and was also significantly faster in providing their responses ($t_1(41)= -2.17$, $p=.036$, $t_2(46)=-3.18$, $p=.003$). The difference in recognition scores for the two Hebb repetition groups showed that including temporal cues resulted in a significantly better recognition of novel items in the by-participants ($t_1(41)=-2.06$, $p=.046$) but not the by-item analysis ($t_2(46)=-1.32$, $p=.195$). There was no difference with regards to RTs between the two Hebb repetition task groups ($t_2(41)= 1.28$, $p=.207$, $t_2(46)= 1.60$, $p=.117$).

In sum, although provision of the temporal grouping cues resulted in a better recognition of novel items, the phoneme monitoring group was still superior in direct recognition of the novel items in comparison to the Hebb repetition group.

Discussion

Experiment 2 tested whether the inclusion of segmentation cues in the Hebb repetition task would support the emergence of lexical integration of novel items. Despite the inclusion of the gaps in the Hebb sequences we did not find any evidence of lexical integration of novel items. In fact, the trend for this comparison was in the opposite direction to that predicted (i.e., facilitation not competition). This result draws into question the generality of the competition effect found by Szmalec et al. (2012). The grouping of the sequences added an extra cue in favour of chunking into trisyllabic wordlike units and increased the explicitness of the task. Encouraging participants to chunk information in a specified manner may have increased task transparency and made participants notice the repetitions. Indeed, analysis of the debriefing questionnaire showed that 14 out of 22 participants (64%) noticed the patterns in syllables and showed awareness as to the task aim. Importantly, Experiment 1 and Experiment 2 provide converging evidence that despite varying the segmentation cues available to participants, the time-course of engagement in lexical competition reported by Szmalec et al. (2012) does not apply in the current circumstances (see supplementary materials for further analysis of awareness and lexical competition).

As Experiment 2 ruled out the possibility that grouping cues are the crucial element of the Hebb repetition task needed to show engagement in lexical competition prior to sleep, the obvious follow up was to test if an increased number of exposures would impact the pattern of lexical competition effects. As stated earlier, we know that increased exposure should lead to lexical competition after a delay including sleep for more explicit training (Gaskell & Dumay, 2003b). Perhaps an equivalent increase in exposure for Hebb repetition will be similarly beneficial. Therefore, in Experiment 3 we tested whether tripling the number of exposures to each novel word (36 presentations) would support the emergence of lexical integration in both training conditions. In this experiment, we also simplified the design of the experiment by eliminating the intermediate 12-hour test condition. Our reasoning was that if the increased number of exposures in Hebb repetition led to competition effects after 24 hours then we could run a further experiment to determine if the effect was also present after 12 hours with or without sleep. However, if the effect was not present after 24 hours then there would be no reason to think that it would emerge after 12 hours.

Experiment 3

Method

In Experiment 3 we tripled the amount of exposure to each novel pseudoword and tested immediately and after 24 hours for the emergence of lexical competition. Both Hebb repetition and phoneme monitoring training methods were used.

Participants. Sixty students from the University of York (forty-six females) participated in this experiment for course credit or financial reward (£6/hour). Their mean age was 20.5 years (ranged from 18 to 31). The criteria for participation were the same as in the previous experiments.

Materials, design and procedures. The critical stimuli were as in the previous experiments. The Hebb repetition task protocol followed that of Experiment 2 in employing grouping cues. This time we increased the number of exposures in both tasks to 36. As mentioned above lexical integration was tested at only two time delays: immediately and 24 hr after encoding. Although the two sessions were always separated by 24 hours, the time of testing itself varied across the day, allowing participants to attend at a wider range of times. Due to the time consuming nature of Hebb repetition training, a simple tripling of the exposure session from Experiment 2 was not feasible in terms of participants' fatigue, as the Hebb repetition training would require over 3 hours to complete. Therefore, we made an adjustment to the ratio of Hebb to filler sequences. Namely, although the Hebb and the Filler sequences were still interleaved, there was only one Filler sequence following two successive but distinct Hebb sequences, each containing different sequences of syllables. Previous studies have demonstrated successful concurrent learning of several different Hebb sequences (Page, Cumming, Norris, McNeil, & Hitch, 2013; Saint-Aubin, Guérard, Fiset, & Losier, 2015). As a result, the order of the presentation of Hebb and Filler trials was: Hebb sequence 1, Hebb sequence 2, Filler sequence. As before there were 4 Hebb sequences in total (three novel words per sequence, so 12 novel words in total), which resulted in 1 hour and 45 minutes to complete the Hebb repetition task. As in Experiment 2, following completion of all experimental tasks, a debriefing questionnaire was administered to determine whether participants became aware of learning novel words and/or the repetition of sequences in the Hebb task. As the phoneme monitoring group was specifically instructed to memorise novel items to increase the explicitness of the training, we expected higher awareness score in this experimental group in comparison to the Hebb repetition group.

Results

For the Hebb repetition training, recall accuracy and regression slopes were calculated according to the previously outlined criteria (see Figure 7). The gradient values for Filler and Hebb trials were significantly different, $F(1,29)=46.33$ $p<.001$, $\eta_p^2=.615$ indicating a higher improvement-gradient for Hebb sequences (M=.009, *SE*=.001) relative to fillers (M=.002, *SE*=.001). Therefore, the Hebb effect was obtained. Inspecting the accuracy scores more closely, it is worth noting that, unlike the previous experiments, there was some evidence that scores were flattening out towards the end of training, suggesting that the extended training had led to participants reaching a ceiling of learning. The fact that this ceiling is well below 100% may in part be a reflection of participants' learning erroneous sequences (cf. Couture, Lafond, & Tremblay, 2008; Lafond, Tremblay, & Parmentier, 2010). The learning curves for the four Hebb sequences did not differ from each other significantly (see Supplementary Materials).

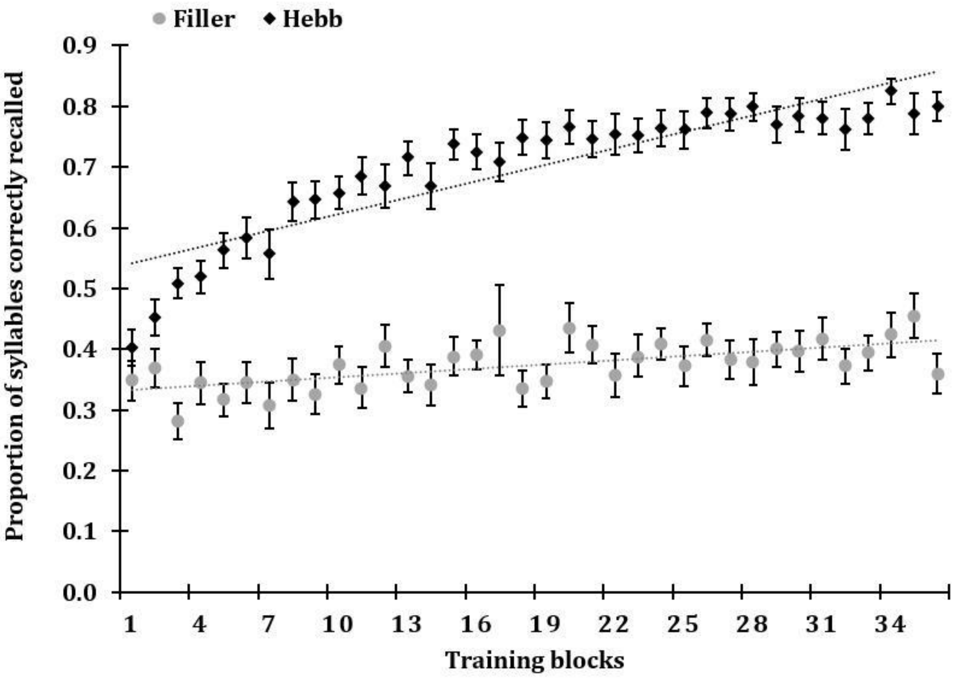


Figure 7. Accuracy (proportion correct) for Hebb and filler sequences in the Hebb repetition task. Values for filler trials represent the average of the two filler sequences presented between each of Hebb sequences (error bars depict standard error; regression lines illustrate the gradient of improvement in performance).

Participant Awareness. In the Hebb repetition group, twenty-one out of thirty participants (70%) reported being aware of the repetition of syllable lists and that they constituted of novel words. Participants' responses listed recognising and learning novel words as an experimental aim. In comparison, in the phoneme monitoring group 97% of all participants stated that learning new words was the aim of the experiment.

Pause detection. Mean RTs and error data are summarised in Table 2. The RT data were analysed using the previously outlined exclusion criteria (both groups ~8%). As in the previous experiment only RTs were analysed for the lexical competition task.

Table 2

Mean Pause Detection Latencies (ms) and Error Percentages for Competitor and Control Conditions in Experiment 3 and Experiment 4.

		Experiment 3				Experiment 4	
Training		Hebb Repetition Task		Phoneme Monitoring		Hebb Repetition Task	
		Competitor	Control	Competitor	Control	Competitor	Control
RT	0-hr	747 (29)	758 (28)	666(29)	681(25)	638 (22)	628 (23)
	24-hr	622 (16)	616 (19)	615 (21)	590(17)	584 (21)	590 (22)
% Err	0-hr	7.8 (1.1)	6.7 (1.0)	6.9 (1.2)	7.8 (1.4)	9.1 (1.6)	8.5 (1.2)
	24-hr	7.5 (1.3)	6.9 (1.1)	6.6 (1.0)	5.3 (9.3)	9.3 (1.5)	7.2 (1.2)

Note. Standard error of the mean in parentheses

After pre-processing as before, the response latencies were entered into a mixed-design ANOVA with the factors Session (0-hr, 24-hr) and condition (Competitor acquisition: competitor versus control), as repeated measures factors, and training task (phoneme monitoring vs. Hebb repetition) as a between-subjects but within-items factor. The analyses revealed a main effect of Session ($F_1(1,58)=49.57$, $p<.001$, $\eta p^2=.461$, $F_2(1,46)=186.17$, $p<.001$, $\eta p^2=.802$), whereas the main effect of training task was nonsignificant in the by-subject analysis (F_1 $(1,58)=2.70$, $p=.106$, $\eta p^2=.044$) but significant in the by-items analysis ($F_2(1,58)=28.26$, $p<.001$, $\eta p^2=.381$). Two interactions were also significant: Session x Competitor acquisition ($F_1(1,58)=8.65$, $p=.005$, $\eta p^2=.013$, $F_2(1,46)=8.65$, $p=.046$, $\eta p^2=.084$) and Session x Training ($F_1(1,58)=4.65$, $p=.035$, $\eta p^2=.074$, $F_2(1,46)=16.46$, $p<.001$, $\eta p^2=.264$). As illustrated in Figure 8, the Session x Competitor acquisition interaction was an indication of a general shift towards stronger lexical competition after 24 hours. Although the Session x Competitor acquisition x Training interaction was nonsignificant ($F_1(1, 58)=1.44$, $p=.235$, $\eta p^2=.024$, $F_2(1,46)=0.88$, $p=.354$, $\eta p^2=.019$), the Session x Training interaction motivated follow-up analyses split by the type of training. For the phoneme monitoring group there was a significant effect of Session ($F_1(1,29)=15.80$, $p<.001$, $\eta_p^2=.353$, $F_2(1,23)=66.94$, $p<.001$, $\eta_p^2=.744$), with response latencies being significantly shorter in Session 2 in comparison to Session 1, and a significant Session x Competitor acquisition interaction ($F_1(1,29)=9.58$, $p=.004$, $\eta_p^2=.248$, $F_2(1,23)=5.74$, $p=.025$, $\eta_p^2=.200$) indicating that the RTs to the test base words became slower in comparison to the control base words (by 24 ms) in the second session that took place 24 hours after the initial learning phase ($F_1(1,29)=5.86$, $p=.022$, $\eta_p^2=.168$, $F_2(1,23)=4.77$, $p=.039$, $\eta_p^2=.172$).

The same analysis for the Hebb group, yielded a significant main effect of Session ($F_1(1,29)=33.98$, $p<.001$, $\eta p^2=.540$, $F_2(1,23)=119.29$, $p<.001$, $\eta p^2=.838$) however the Session x Competitor acquisition interaction was non-significant ($F_1(1,29)=1.37$, $p=.251$, $\eta p^2=.045$, $F_2(1,23)=.51$, $p=.482$, $\eta p^2=.022$), and there was no significant competition effect after a 24 hour delay (6 ms difference in RTs to test and control base words; $F_1(1,29)=.45$, $p=.508$, $\eta p^2=.015$, $F_2(1,23)=.37$, $p=.550$, $\eta p^2=.016$). Therefore, it appears that the shift towards stronger lexical competition after a consolidation period was driven largely by the phoneme monitoring training.

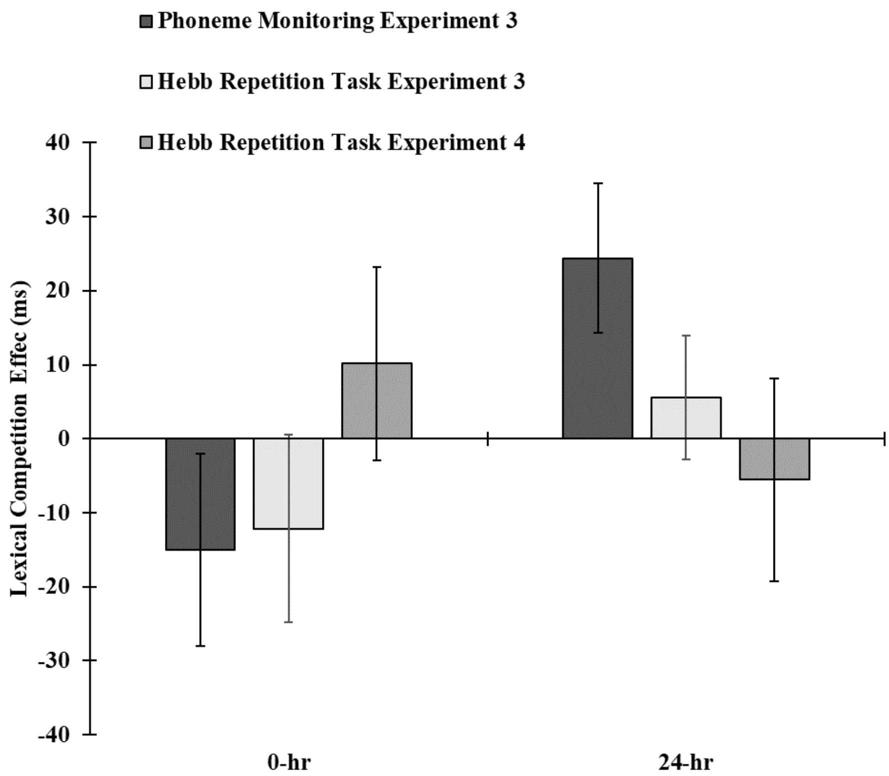


Figure 8. Lexical competition effect (competitor RT- control RT) across two sessions for phoneme monitoring and Hebb repetition groups. Error bars represent standard error of the means and are not adjusted to facilitate within-participants comparisons, given the mixed design (Cousineau & Brien, 2014).

Cued recall and 2AFC. The responses in the cued recall task were scored as in Experiment 1 and 2. The increased number of presentations of novel pseudowords resulted in higher recall in both groups in comparison to the previous experiment where less than 30% of items were recalled. The Hebb repetition group recalled 43% of novel items and the phoneme monitoring group significantly more (58%; $t_1(58)= 2.96$, $p=.004$, $t_2(46)=2.79$, $p=.008$). Similarly, in the 2AFC task both groups scored above the chance level (Hebb repetition task: $t_1(29)=18.93$, $p<.001$, $t_2(23)=9.53$, $p<.001$; Phoneme monitoring: $t_1(29)=33.94$, p<.001, $t_2(23)=28.67$, $p<.001$), with the phoneme monitoring group recognising significantly more items (93% vs. 84%) ($t_1(58)=3.88$, $p<.001$, $t_2(46)=2.08$, $p=.043$) with shorter RTs ($t_1(58)=-4.23$, $p<.001$, $t_2(46)=-5.89$, $p<.001$). The results of the explicit tests are illustrated in Figure 9.

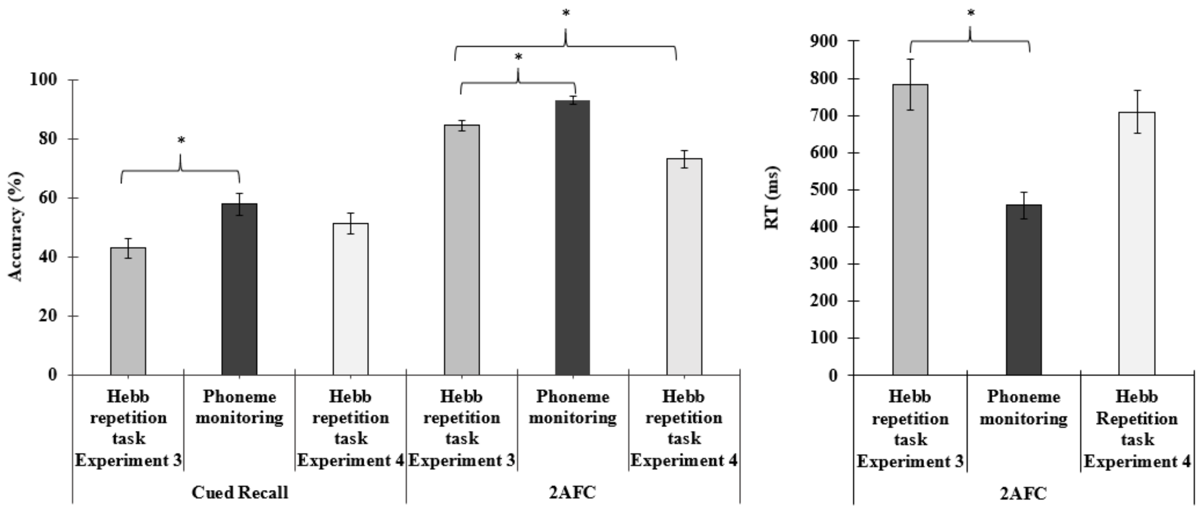


Figure 9. Mean accuracy in explicit tests for the Hebb repetition task and phoneme monitoring groups and mean RTs for both experimental groups in 2AFC task. Error bars represent standard error of the means.

Discussion

Experiment 3 investigated whether an increased number of exposures would lead to a better encoding of the novel pseudowords and aid lexical integration of novel items. Unlike previous experiments, we found a change in the lexical competition profile over time, with stronger competition after a day than immediately after encoding. Although there was no three-way interaction in the analyses, an interaction between type of training and session suggested differences in the effect of time for the two types of encoding. When the training methods were tested separately, there was evidence of lexical competition emerging only after a delay for phoneme monitoring, but no similar evidence for Hebb repetition.

The extended encoding session in the Hebb repetition condition allowed 12 syllable sequences to be encountered 36 times each over a period of almost 2 hours. Despite this high level of exposure (three times that used by Szmalec et al., 2012) there was no evidence of these sequences engaging in lexical competition immediately after or 24 hours later. On the other hand, and in contrast to Experiment 1, we found a significant lexical competition effect after a 24 hour delay in the phoneme monitoring condition. This suggests that a good level of encoding and a consolidation delay that contains sleep are beneficial for lexical integration of new items learned explicitly, which stands in agreement with previous studies on novel word learning (Bakker et al., 2014; Davis et al., 2009; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003b; Henderson et al., 2012).

It is possible that the observed results are not so much dependent on the different learning mechanisms utilised by the two groups, but are more a consequence of the fact that participants never encountered the new items as whole words in the Hebb task, and so the acoustic mismatch between the isolated syllables of the novel item (e.g., "*bih-kee*") and the onset of the contiguous existing word (e.g., *bikini*) was too great to influence lexical competition. This is indeed quite feasible, but it is worth noting that part of the argument underlying the Hebb task as a model of learning is that chunking will automatically and implicitly generate continuous "word" sequences. Indeed, the original study by Szmalec et al. (2012) demonstrated effects of lexical competition for isolated syllables that were presented in written form, which clearly have even less overlap with the contiguous spoken word sequences. Therefore, this cannot be the whole story. Furthermore, the performance of the Hebb group when asked to explicitly recall the syllable sequences was reasonably good (43%) and their ability to pick out these sequences from foil sequences when presented with the syllables contiguously was even better (84% correct). Based on the debriefing questionnaire, 70% of the Hebb group reported to be aware that separate strings embedded in the Hebb sequences consisted of novel words. The debriefing questionnaire results together with improved performance in the cued recall and 2AFC tasks suggest that it is unlikely that participants did not extract the novel syllable sequences in any form (see supplementary materials for analyses of the pause detection data for correctly recalled items in the cued recall task).

A plausible cause of the difference in Hebb repetition effects between studies pertains to the relationship between the novel and existing words. As mentioned, similar to previous explicit word-learning studies, we used novel items that were fairly distinct neighbours of their English counterparts (i.e. deviating in the full final syllable) as opposed to the Szmalec and colleagues Hebb repetition studies, which used items that more closely overlapped with their English base words (i.e. only the final vowel deviation). In doing this, we wanted to test whether any effects of the Hebb repetition procedure would extend to competition neighbourhoods more generally. Perhaps then the minimally deviant pseudowords used by Szmalec and colleagues in their studies actually activated the neighbouring words (Cutler, Sebastián-Gallés, Soler-Vilageliu, & van Ooijen, 2000; van Ooijen, 1996) in a way that led to the novel word being treated as matching the existing word, perhaps as a new phonological variant (Bürki & Gaskell, 2012). This automatic activation would not occur in our case because of the more substantial mismatch between novel and existing word.

Evidence consistent with this explanation comes from a study of novel word learning in French. Over 4 days, Bürki, Spinelli, and Gaskell (2012) taught participants novel monosyllabic spoken forms that could potentially be reduced forms of a bisyllabic word (e.g., participants learned “plour”, which might be a reduced form of “pelour”). Later disambiguating orthographic information (the words were spelt out either as bisyllables or monosyllables) had knock-on effects in terms of both the pronunciation and recognition of the spoken forms. Importantly, this new information did not show any influence of consolidation over 24 hours. Thus, the nature of the learning experience and the similarity of the new form to an existing form can shape the need for consolidation. It is important to note that Bürki et al. (2012) studied the effect of an additional vowel rather than a different vowel, and their participants had existing knowledge of similar vowel reductions (i.e, the reduction was reasonably systematic), whereas Szmalec and colleagues (2012) used more irregular deviations (e.g., *bikina/bikini*). Nonetheless, the Bürki et al. (2012) study and similar studies of regular variants (Snoeren, Gaskell, Maria, & Di Betta, 2009) add weight to the argument that single vowel deviations from existing words might rely less on consolidation than more substantial deviations. We tested this possibility in Experiment 4 where we addressed the degree of mismatch between the novel pseudowords and their English counterparts.

Experiment 4

Method

In Experiment 4 we altered the novel sequences so that they only deviated from the existing words in the final vowel (e.g., for the existing word *bikini*, instead of *bikiso* we used *bikino*) and tested whether Hebb repetition learning of these sequences would lead to engagement in lexical competition. We hypothesised, based on the difference between our results and Szmalec et al (2012), that if a higher level of overlap between novel pseudowords

and existing English words is the crucial factor in the lexicalisation process following Hebb-style learning, we would observe a lexical competition effect after a 24-hour delay.
Participants. Thirty students from the University of York (twenty-six females) participated in this experiment for course credit or financial reward (£6/hour). Their mean age was 20.4 years (ranged from 18 to 31). The criteria for participation were the same as in the previous experiments.

Materials, design and procedure. The critical stimuli were 24 pseudowords, similar to the ones used in the previous experiments, with the exception that this time they differed from their English counterparts in the final vowel only (see Appendix A for a complete list). The pseudowords were recorded using the same speaker and procedure as previous experiments. The Hebb repetition task protocol followed Experiment 3 in employing grouping cues, a high level of exposure (36 presentations of each novel word), and two test points: immediately and 24 hr after encoding.

Results

For the Hebb repetition training, recall accuracy and regression slopes were calculated according to the previously outlined criteria (see Figure 10). The gradient values for Filler and Hebb trials were significantly different, $F(1,29)=32.61$ $p<.001$, $\eta p^2=.529$ indicating a steeper improvement-gradient for Hebb sequences (M=.007, SE=.001) relative to fillers (M=.002, SE=.001). As in Experiment 3, there was some evidence that scores were flattening out towards the end of training. The four separate Hebb sequences were learned at a similar level (see Supplementary Materials).

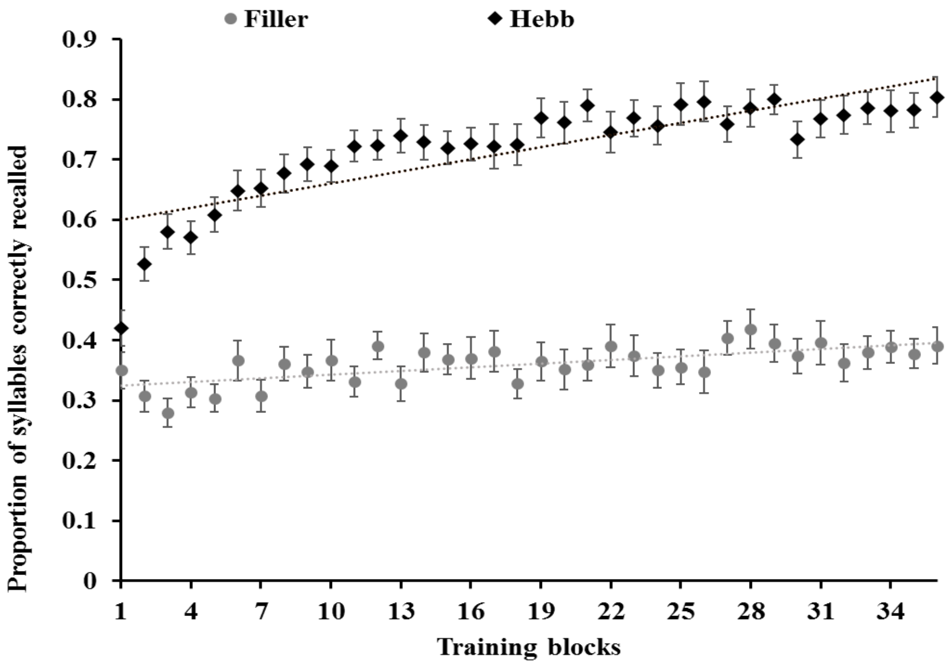


Figure 10. Accuracy (proportion correct) for Hebb and filler sequences in the Hebb repetition task. Values for filler trials represent the average of the two filler sequences presented between each of Hebb sequences (error bars depict standard error; regression lines illustrate the gradient of improvement in performance).

Participant Awareness. Twenty-two out of thirty participants (73%) reported being aware of the repetition of syllable lists and that they constituted of novel words. Participants' responses listed recognising and learning novel words as an experimental aim.

Pause detection. Mean RTs and error data are summarised in Table 2. The RT data were analysed. As in Experiment 3, only RTs were analysed using the previously outlined exclusion criteria (~8% of data points excluded). These were entered into a mixed-design ANOVA with the repeated-measures independent variables Session (0-hr, 24-hr) and Competitor acquisition (competitor versus control). The analyses revealed a main effect of Session ($F1(1,29)=10.30$, $p=.003$, $\eta p^2=.262$, $F2(1,23)=23.31$, $p<.001$, $\eta p^2=.503$), whereas the main effect of Competitor was not significant ($F1\ (1,29)=.07$, $p=.790$, $\eta p^2=.002$; $F2(1,23)=.01$, $p=.915$, $\eta p^2=.001$). The Session x Competitor acquisition interaction was also not significant ($F1(1,29)=.58$, $p=.455$, $\eta p^2=.019$; $F2(1,23)=.58$, $p=.453$, $\eta p^2=.025$) indicating that the RTs to the test base words were not significantly different from the control base words. In fact, RTs to the base words and control words were almost identical in the second session that took place 24 hours after the initial learning phase (5 ms difference in the wrong direction for lexical competition; see Figure 8).

Cued recall and 2AFC. The responses in the cued recall task were scored as in Experiments 1-3 (see Figure 9). The deviation of just the final vowel, as compared to deviation of final syllable, between novel words and their existing counterparts resulted in slightly higher recollection of novel items in comparison to the previous experiment (51% versus 43% to the Hebb repetition group in Experiment 3), but this difference was not significant ($t_1(58)= 1.67$, $p=.101$; $t_2(46)=1.63$, $p=.111$). In the 2AFC participants scored above chance level (mean accuracy 73%; $t_1(29)=18.77$, $p<.001$, $t_2(23)=23.12$, $p<.001$). Participants in Experiment 4 showed poorer recognition of novel items in comparison to participants in Experiment 3 (73% versus 84%, respectively; see Figure 9). This difference was significant ($t_1(47.45)=-3.19$, $p=.003$, $t_2(46)=-2.22$, $p<.031$) but the groups had similar RTs ($t_1(58)=-.83$, $p=.411$, $t_2(46)=45.75$, $p<.204$).

Discussion

Experiment 4 investigated whether an increased similarity between novel items and existing English words (i.e. deviation of final vowel only as opposed to deviation of final syllable) would aid their lexical integration when encoded via Hebb-style learning. Despite a substantial similarity between novel pseudowords and existing English words (similar to Szmalec et al., 2012) we did not observe any evidence of a lexical competition effect at any time-point. This indicates that the novel pseudowords were not consolidated sufficiently to engage in lexical competition with the existing words. Moreover, although cued recall performance was numerically higher in Experiment 4 in comparison to Experiment 3 (51% versus 43%), accuracy on the 2AFC was in fact significantly lower (73% versus 84%). This suggests that the increased similarity between novel items and existing English words, which made the novel pseudowords less distinct, hindered participants' ability to discriminate them from their near neighbours. These results indicate that the level of mismatch between novel pseudowords and existing real words may not influence the time course of lexical integration of novel items.

General Discussion

The research presented here is the first attempt to evaluate the lexical impact of two different approaches to word learning by comparing a largely explicit form of training utilising phoneme monitoring with a more implicit Hebb repetition learning paradigm. In a series of four experiments we tested whether the Hebb repetition procedure would facilitate the time course of lexical integration of novel words compared with a more explicit phoneme monitoring task. Considering just Hebb repetition conditions, in Experiment 1, as is typical in Hebb studies, we did not provide any temporal grouping cues to boundary locations, with only statistical information marking the potential word boundaries. In Experiment 2 we supplemented the statistical cues with temporal cues to word boundaries following Szmalec and colleagues (2012). In Experiment 3 we employed the same temporal and statistical cues but tripled the length of the training session. Finally, In Experiment 4 we explored whether the level of mismatch between novel words and their real word counterparts affects the lexical integration process after the Hebb repetition task. In all four experiments we found no evidence that Hebb-style learning leads to accelerated integration of novel items prior to sleep. In fact, even after sleep we found no evidence of Hebb repetition leading to heightened competition between novel and existing words.

Several Hebb repetition studies have reported lexical integration of novel words, (Bogaerts, Szmalec, Hachmann, Page, & Duyck, 2015; Szmalec et al., 2012) so we have evidence that this effect can be found in some circumstances. Perhaps then our null effects are simply unfortunate failures to reach significance levels for a real but not substantial underlying effect. In order to test this possibility with enhanced statistical power, we ran a meta-analysis of all the Hebb repetition learning conditions in our four experiments. Based on Szmalec et al. (2012), a lexical competition effect should be present after 12 hours or more regardless of whether the delay between learning and testing contained sleep. Hence, in our experiments it should be observed after both 12 and 24-hour delays in Experiments 1 and 2 and after 24 hours in Experiments 3 and 4. Therefore, we analysed the Hebb condition pause detection competition effects from these six conditions. Two of the four conditions showed a numerical difference in the predicted direction (13 ms, 6 ms) and four showed a difference in the non-predicted direction (-18 ms, -10 ms, -7 ms, -6 ms). Overall the difference was in the nonpredicted direction (-3 ms) and was not significant ($F1(1, 140) = .39$, $p = .54$, $\eta p^2=.003$; $F2(1, 23) = .70$, $p = .413$, $\eta p^2=.029$).

To check the informativeness of this null result, we computed the Bayes Factor (Dienes, 2014) for the overall Hebb effect of -3 ms in comparison with the effect for more explicit training found after a delay given sufficient exposure in Experiment 3 (24 ms). The Bayes Factor allows statistical assessment of the strength of evidence for or against a null hypothesis, with a value of 3 or more indicating substantial evidence against the null hypothesis and of 1/3 or less as evidence for the null hypothesis. The Bayes Factor was calculated according to Dienes (2008), resulting in a value of 0.16 based on the participants analysis and 0.18 based on the items analysis. Thus our data do provide substantial evidence for the null hypothesis that Hebb repetition in our study did not induce lexical competition after delays of 12-24 hours.

The four experiments explored a range of points in the parameter space of variables that could impact the learning process and the emergence of lexical competition for Hebb repetition, such as grouping cues in the Hebb repetition task, the level of exposure, the amount of time available for consolidation and the degree of overlap between novel and real words. It is reasonable to assume that the changes in all these factors sampled a range of levels of implicitness of the task, but at no point did competition effects emerge for the Hebb condition. It is clear that the level of implicitness of the Hebb task depends on quite a few factors, and although it is generally described as implicit in nature (e.g., Smalle et al., 2017), in some of the experiments here in fact a majority of the participants became aware of the nature of the task. In additional analyses we assessed the serial position curves in the Hebb task across our four experiments and observed an altered shape across the series following the insertion of temporal pauses during list presentation (see Supplementary Materials). In exploratory analyses we also assessed whether participants' awareness has any impact on the emergence of lexical competition but found no effect (see Supplementary Materials).

In contrast, the level of exposure was important for the more explicit form of learning, with greater exposure leading to lexical competition after a 24 hour delay in Experiment 3. Similarly, the explicit measures of memory for novel items indicated better performance after the explicit training in comparison to the Hebb task. Consistent with previous studies (Bakker et al., 2014; Davis et al., 2009; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003b; Henderson et al., 2012), this suggests that the emergence of lexical competition is optimized by two factors: a good level of initial explicit encoding and a time delay that includes sleep. As described in the introduction, this characterization fits well with a complementary systems account of word learning (Davis & Gaskell, 2009). We remain very open to the possibility or even likelihood that such an account is too simplistic to explain all aspects of word learning: there are likely to be other neural mechanisms that come into play in certain circumstances (cf. McMurray et al., 2016). Nonetheless, we do not find any evidence in this study that the implicit mechanisms that underlie Hebb repetition can lead to similar engagement in lexical competition.

A recent study of predictors of language attainment also suggested that explicit measures of memory are of more relevance to language outcome measures than implicit measures. West, Vadillo, Shanks, and Hulme (2018) tested a set of 7-8-year-old children on a large battery of explicit and implicit memory tests to determine which were predictive of good language and literacy attainment. Their results showed strong correlations between the explicit declarative memory tests and attainment (e.g., word list learning), but very weak correlations for the implicit tests. Interestingly, explicit immediate serial recall performance—as used in the Hebb repetition task—was a good predictor of language attainment, but the implicit gain attributed to Hebb repetition was a poor predictor. Although cross-sectional, this study casts doubt on the proposal that implicit learning skills are crucial to language learning and may underlie some language learning disorders (Ullman, 2004; (although see also Mosse & Jarrold, 2008).

Given that the failure of Hebb repetition to induce engagement in lexical competition is robust in our study, and that Szmalec and colleagues have found this effect in several studies, we need to consider what might be the factor that dictates the presence or absence of this effect. There were several differences between the two sets of studies so we cannot be certain at this point. One possibility is the number of words to be learnt. Szmalec et al., (2012) used six novel pseudowords, whereas in our study we used twice as many. Possibly we overloaded the mechanism by which new words are learned, although robust Hebb effects were found in all our experiments, indicating learning of the sequences, and 2AFC recognition of the form of the novel words in this condition was reasonable (above 70% in Experiment 2 and 4; above 80% in Experiment 3). Clearly Hebb repetition was leading to knowledge of the appropriate chunks in our experiments, but this was not sufficient to influence recognition of neighbouring existing words. Another difference that does not seem likely to be influential is the modality of presentation. Szmalec and colleagues used written syllables (in Dutch), but for English these would have been too ambiguous in pronunciation and so we opted for spoken syllables to ensure that the correct vowels were learned. But the use of spoken syllables would seem to enhance the likelihood of competition in the auditory modality, given that Bakker et al. (2014) found that transfer from written word learning to engagement in auditory lexical competition is delayed compared with the opposite transfer or intramodal effects.

One intriguing possibility relates to the composition of the filler sequences in the Hebb task.[1] Several studies have indicated that learning via the Hebb repetition task is more robust if the Hebb and filler sequences are constructed from non-overlapping items (Page, et al.,2013; Johnson, Dygacz, and Miles, 2017). For example. Page, et al. (2013) showed that learning was weakened or abolished when Hebb and filler sequences were all permutations of the same items (in this case, monosyllabic nouns). Szmalec et al. (2012) opted for reusing the same pool of syllables for both Hebb and filler lists in their study. However, given the subsequent evidence of Page et al. (2013), we used non-overlapping syllables for fillers and Hebb sequences in our four experiments in order to maximise the opportunity for Hebb learning and hence, we assumed, aid lexicalisation of the Hebb triplets.

Intriguingly, recent findings by Smalle et al. (2017) suggest that different neural mechanisms may operate when Hebb sequences contain overlapping versus non-overlapping fillers are used. The authors wanted to investigate why adults are disadvantaged compared with children in the learning of non-overlapping (but not overlapping) Hebb sequences. Their hypothesis was that executive functions and contributors to declarative memory sited in the dorsolateral prefrontal cortex were inhibiting the procedural learning of the non-overlapping sequences in adults. They found that disruption of this area via Transcranial Magnetic Stimulation impaired the Hebb learning of non-overlapping sequences in young adults. This disruption was not the case for overlapping sequences. The authors suggested that overlap between Hebb and filler sequences "seems to counteract the sequential learning processes that underlie Hebb repetition learning", consequently recruiting "different declarative-based memory resources, or more attentional control" (Smalle et al., 2017, p. 4).

If the above dissociation is correct then there is a neat, if speculative, way to explain both our evidence of no lexical integration of novel words learned using non-overlapping sequences, and Szmalec et al.'s (2012) positive evidence for lexical integration using overlapping sequences. When Hebb repetition learning engages more procedural neural circuitry (Smalle et al., 2017) for sequences that do not overlap with the filler sequences then we have robust evidence in the four current experiments for the learned sequences failing to engage in lexical competition with existing results. However, when this procedural learning is disrupted by the more challenging task of learning transitional probabilities in overlapping sequences, declarative-based memory resources are engaged, and Hebb repetition learning leads to lexical competition. Consistent with this dissociation, although Hebb repetition learning is typically viewed as independent of the hippocampus (e.g., Gagnon et al., 2004), Kalm, Davis & Norris (2013) found evidence for hippocampal involvement in multiple overlapping sequences, again pointing to the recruitment of a second neural system depending on the complexity of the mapping. Interestingly, a parallel debate is emerging in terms of the involvement of the hippocampus in "fast mapping" learning of words, with initial results suggesting hippocampus independent learning (Sharon, Moscovitch, & Gilboa, 2011), but the more recent balance of evidence favouring a crucial role for the hippocampus (Cooper et al.,

[1] We thank an anonymous reviewer for this suggestion.

2018). If the above speculative account turns out to be correct, then the argument that Hebb repetition learning provides a good model of vocabulary learning needs at minimum some qualification in that a pure procedural model appears insufficient.

Conclusions

Lexical integration of novel words was tested in four experiments using the Hebb repetition task as an example of implicit statistical learning and phoneme monitoring as a more explicit means of familiarisation. We observed evidence for engagement of the novel words in lexical competition only for the more explicitly trained words, and only when the initial exposure level was high. Successful lexical integration of novel items appears to benefit from a sufficient level of explicit exposure followed by a consolidation opportunity that includes sleep. Our findings do not provide evidence for the implicit mechanisms underlying Hebb repetition as effective for learning and, particularly, integration of verbal material. While we do not doubt the value of implicit and statistical learning mechanisms for language learning more generally, it appears that explicit memory systems play a crucial role in acquiring and retaining information about word forms. Discrepancies between our findings and previous studies of the Hebb repetition effect may be a consequence of modality of presentation and/or composition of the filler material. When novel words are learnt well the requirement for consolidation may be reduced, but for the more general process of acquiring lexical neighbours offline consolidation appears to be a crucial part of the process.

Acknowledgements

This research was funded by the Economic and Social Research Council studentship awarded to JMS, and an ESRC grant (ES/N009924/1) awarded to Lisa Henderson and Gareth Gaskell. The authors thank Mike Page for his helpful comments on design and interpretation, and Anna á Váli Guttesen and Ismail Abdelrahim for their help with data collection. The authors declare that there are no conflicts of interest.

References

Baayen, R., Piepenbrock, R., & van Rijn, H. (1993). The {CELEX} lexical data base on {CD-ROM}. Retrieved from http://www.citeulike.org/group/1778/article/930018

Baddeley, A. D., & Warrington, E. K. (1970). Amnesia and the distinction between long-and short-term memory. Journal of Verbal Learning and Verbal, 9(2), 176–189.

Bakker, I., Takashima, A., Hell, J. G. Van, & Janzen, G. (2015). Changes in Theta and Beta Oscillations as Signatures of Novel Word Consolidation. *Journal of Cognitive Neuroscience*, (March), 1286–1297. https://doi.org/10.1162/jocn

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & Mcqueen, J. M. (2015). Changes in Theta and Beta Oscillations as Signatures of Novel Word Consolidation. *Journal of Cognitive Neuroscience*, 1286–1297. https://doi.org/10.1162/jocn.

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Competition from unseen or unheard novel words: Lexical consolidation across modalities. *Journal of Memory and Language*, *73*, 116–130. https://doi.org/10.1016/j.jml.2014.03.002.

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English Lexicon Project. Behavior Research Methods, 39, 445-459.

Batterink, L. J., Reber, P. J., & Paller, K. A. (2015). Functional differences between statistical learning with and without explicit training. *Learning & Memory*, *22*(November), 544–556. https://doi.org/10.1101/lm.037986.114

Bayley, P. J., Reilly, R. C. O., Curran, T., & Squire, L. R. (2008). New Semantic Learning in Patients With Large Medial Temporal Lobe Lesions. *Hippocampus*, *18*, 575–583. https://doi.org/10.1002/hipo.20417

Bogaerts, L., Szmalec, A., Hachmann, W. M., Page, M. P. A., & Duyck, W. (2015). Linking memory and language: Evidence for a serial-order learning impairment in dyslexia. *Research in Developmental Disabilities*, *43–44*, 106–122. https://doi.org/10.1016/j.ridd.2015.06.012Born, J., & Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychological Research*, *76*(May), 192–203. https://doi.org/10.1007/s00426-011-0335-6

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours : The impact of novel word learning on the identification of visually similar words. *Cognition*, *97*, 45–54. https://doi.org/10.1016/j.cognition.2005.02.002

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English. *Behavior Research Methods*, *41*(4), 977–990.

Bürki, A., & Gaskell, M. G. (2012). Lexical Representation of Schwa Words : Two Mackerels , But Only One Salami. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(3), 617–631. https://doi.org/10.1037/a0026167

Bürki, A., Spinelli, E., & Gaskell, M. G. (2012). A written word is worth a thousand spoken words: The influence of spelling on spoken-word production. *Journal of Memory and Language*, *67*, 449–467. https://doi.org/10.1016/j.jml.2012.08.001

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. In *Proceedings of the Stanford Child Language Development, 15,* (pp. 17–29).

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.

Cooper, E., Greve, A., & Henson, R. N. (2018). Little evidence for Fast Mapping (FM) in adults: A review and discussion. Cognitive neuroscience, 1-14.

Cousineau, D., & Brien, F. O. (2014). Error bars in within-subject designs : a comment on Baguley (2012), 1149–1151. https://doi.org/10.3758/s13428-013-0441-z

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Fast Mapping Rapidly Integrates Information Into Existing Memory Networks, *143*(6), 2296–2303.

Couture, M., & Tremblay, S. (2006). Exploring the characteristics of the visuospatial Hebb repetition effect. *Memory & Cognition*, *34*(8), 1720–1729. https://doi.org/10.3758/BF03195933

Couture, M., Lafond, D., & Tremblay, S. (2008). Learning correct responses and errors in the Hebb repetition effect: Two faces of the same coin. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(3), 524-532.

Cumming, N., Page, M., & Norris, D. (2003). Testing a positional model of the Hebb effect. *Memory*, *11*(1), 43–63. https://doi.org/10.1080/741938175

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: cross-linguistic comparisons. *Memory & Cognition*, *28*(5), 746–755. https://doi.org/10.3758/BF03198409

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, *21*, 803– 820. https://doi.org/10.1162/jocn.2009.21059

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(0), 3773–3800. https://doi.org/10.1098/rstb.2009.0111

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews. Neuroscience*, *11*(2), 114–126. https://doi.org/10.1038/nrn2762

Dienes, Z. (2008). *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference*. Palgrave Macmillan.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*(781), 1–17. https://doi.org/10.3389/fpsyg.2014.00781

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words: Research report. *Psychological Science*, *18*, 35–39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

Dumay, N., Gaskell, M. G., & Feng, X. (2004). A day in the life of a spoken word. In *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 339–344). Mahwah, NJ: Lawrence Erlbaum Associates: Mahwah, NJ: Erlbaum.

Duyck, W., Desmet, T., Verbeke, L., & Brysbaert, M. (2004). WordGen: A Tool for Word Selection and Non-Word Generation in Dutch, German, English, and French. Behavior Research Methods, Instruments & Computers, 36(3), 488-499.

Fernandes, T., Kolinsky, R., & Ventura, P. (2009). The metamorphosis of the statistical
segmentation output: Lexicalization during artificial language learning. *Cognition*, *112*(3), 349–366. https://doi.org/10.1016/j.cognition.2009.05.002

Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *PNAS*, *103*(31), 11778–11783.

Forster, K. I., & Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *35*(1), 116–124. https://doi.org/10.3758/BF03195503

Gagnon, S., Foster, J., Turcotte, J. J., & Jongenelis, S. (2004). Involvement of the hippocampus in implicit learning of supra-span sequences: The case of SJ. *Cognitive Neuropsychology*, *21*(8), 967–882. https://doi.org/10.1080/02643290342000609

Gaskell, M. G., & Dumay, N. (2003a). Effects of vocabulary acquisition on lexical competition in speech perception and production. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th ICPhS Conference* (pp. 1485–1488). Adelaide, Australia: Causal Productions.

Gaskell, M. G., & Dumay, N. (2003b). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105–132. https://doi.org/10.1016/S0010-0277(03)00070-2

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition inthe perception of spoken words. *Cognitive Psychology*, *45*(August 1997), 220–266. https://doi.org/10.1016/S0010-0285(02)00003-8

Geukes, S., Gaskell, M. G., & Zwitserlood, P. (2015). Stroop effects from newly learned color words: Effects of memory consolidation and episodic context. *Frontiers in Psychology*, *6*(MAR), 1–16. https://doi.org/10.3389/fpsyg.2015.00278

Greve, A., Cooper, E., & Henson, R. N. (2014). No evidence that “fast-mapping” benefits novel learning in healthy Older adults. *Neuropsychologia*, *60*, 52–59. https://doi.org/10.1016/j.neuropsychologia.2014.05.011

Guérard, K., Saint-Aubin, J., Boucher, P., & Tremblay, S. (2011). The role of awareness in anticipation and recall performance in the Hebb repetition paradigm: implications for sequence learning. *Memory & Cognition*, *39*, 1012–1022. https://doi.org/10.3758/s13421-011-0084-1

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. *Brain Mechanisms and Learning: A Symposium*, 37–46.

Henderson, L. M., Devine, K., Weighall, A. R., & Gaskell, M. G. (2015). When the daffodat flew to the intergalactic zoo: Off-line consolidation is critical for word learning from stories. *Developmental Psychology*, *51*(3), 406–417. https://doi.org/10.1037/a0038786

Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, M. G. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science*, *15*, 674– 687. https://doi.org/10.1111/j.1467-7687.2012.01172.x

Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, M. G. (2013). Online lexical competition during spoken word recognition and word learning in children and adults. *Child Development*, *84*(5), 1668–1685. https://doi.org/10.1111/cdev.12067

Horst, J. S., & Samuelson, L. K. (2008). Fast Mapping but Poor Retention by 24-Month-Old Infants. *Psychology Press*, *13*(2), 128–157. https://doi.org/10.1080/1525O(xH)701795598

Horton, N., Hay, D. C., & Smyth, M. M. (2008). Hebb repetition effects in visual memory : The roles of verbal rehearsal and distinctiveness. *The Quarterly Journal of Experimental Psychology*, *61*(12), 1769–1777. https://doi.org/10.1080/17470210802168674

Johnson, A. J., Cauchi, L., & Miles, C. (2013). Hebbian learning for olfactory sequences. *The Quarterly Journal of Experimental Psychology*, *66*(6), 1082–1089. https://doi.org/10.1080/17470218.2012.729068

Johnson, A. J., Dygacz, A., & Miles, C. (2017). Hebb repetition effects for non-verbal visual sequences: Determinants of sequence acquisition. Memory, 25(9), 1279-1293.

Johnson, A. J., Shaw, J., & Miles, C. (2016). Tactile order memory : evidence for sequence learning phenomena found with other stimulus types. *Journal of Cognitive Psychology*, *28*(6), 718–725. https://doi.org/10.1080/20445911.2016.1186676

Kalm, K., Davis, M. H., & Norris, D. (2013). Individual sequence representations in the medial temporal lobe. Journal of cognitive neuroscience, 25(7), 1111-1121.

Kapnoula, E. C., Gupta, P., Packard, S., & McMurray, B. (2015). Immediate lexical integration of novel word forms. *Cognition*, *134*, 85–99. https://doi.org/10.1016/j.cognition.2014.09.007

Kapnoula, E. C., & McMurray, B. (2015). Newly learned word forms are abstract and integrated immediately after acquisition. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-015-0897-1

Lafond, D., Tremblay, S., & Parmentier, F. (2010). The ubiquitous nature of the Hebb repetition effect: Error learning mistaken for the absence of sequence learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36(2), 515-522.

Lindsay, S., & Gaskell, M. G. (2009). Spaced learning and the lexical integration of novel words. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2517–2522). Austin, TX: Cognitive Science Society.

Mattys, S. L., & Clark, J. H. (2002). Lexical activity in speech processing: Evidence from pause detection. *Journal of Memory and Language*, *47*, 343–359. https://doi.org/10.1016/S0749-596X(02)00037-2

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–57.

McMurray, B., Kapnoula, E. C., & Gaskell, M. G. (2016). Learning and integration of new word-forms: Consolidation, pruning and the emergence of automaticity. In M. G. Gaskell & J. Mirković (Eds.), *Speech Perception and Spoken Word*.

Mosse, E. K., & Jarrold, C. (2008). Hebb learning, verbal short-term memory, and the acquisition of phonological forms in children. *Quarterly Journal of Experimental Psychology (2006)*, *61*(July 2014), 505–514. https://doi.org/10.1080/17470210701680779

Nazzi, T., Gopnik, A., & Karmiloff-Smith, A. (2005). Asynchrony in the cognitive and lexical development of young children with Williams syndrome. *Journal of Child Language*, *32*(2), 427–438. https://doi.org/10.1017/S0305000904006737

Nespor, M., Peña, M., & Mehler, J. (2003). On the Different Roles of Vowels and Consonants in Speech Processing and Language Acquisition. *Lingue E Linguaggio*, (2), 203–230. https://doi.org/10.1418/10879

Norris, D. (1994). Shortlist : a connectionist model continuous speech recognition. *Cognition*, *52*, 189–234.

Page, M. P. A., Cumming, N., Norris, D., Hitch, G. J., & McNeil, A. M. (2006). *Repetitionlearning in the immediate serial recall of visual and auditory materials. Journal of Experimental Psychology:learning, Memory and Cognition* (Vol. 32).

Page, M. P. A., Cumming, N., Norris, D., McNeil, A. M., & Hitch, G. J. (2013). Repetition-spacing and item-overlap effects in the Hebb repetition task. *Journal of Memory and Language*, *69*(4), 506–526. https://doi.org/10.1016/j.jml.2013.07.001

Page, M. P., & Norris, D. (2008). Is there a common mechanism underlying word-form learning and the Hebb repetition effect? Experimental data and a modelling framework. In *Interactions Between Short-Term and Long-Term Memory in the Verbal Domain.* (pp. 136–156).

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child Development*, *80*(3), 674–685.

Rasch, B., & Born, J. (2013). About Sleep's Role in Memory. *Physiol Rev*, *93*, 681–766. https://doi.org/10.1152/physrev.00032.2012

Saffran, J. R. (2002). Constraints on Statistical Language Learning. *Journal of Memory and Language*, *196*(47), 172–196. https://doi.org/10.1006/jmla.2001.2839

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*, 110–114. https://doi.org/10.1111/1467-8721.01243

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*, 8–11.

Saffran, J. R., Senghas, A., & Trueswell, J. C. (2000). The acquisition of language by children. *PNAS*, *98*(23), 12874–12875.

Saint-Aubin, J., Guérard, K., Fiset, S., & Losier, M. (2015). Learning Multiple Lists at the Same Time in the Hebb Repetition Effect. *Journal of Experimental Psychology*, *69*(1), 89–94.

Schwartz, M., & Bryden, M. P. (1971). Coding factors in the learning of repeated digit sequences. Journal of Experimental Psychology, 87(3), 331±334.

Sharon, T., Moscovitch, M., & Gilboa, A. (2011). Rapid neocortical acquisition of long-term arbitrary associations independent of the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 1146–1151. https://doi.org/10.1073/pnas.1005238108

Smalle, E. H. M., Panouilleres, M., Szmalec, A., & Möttönen, R. (2017). Language learning in the adult brain: disrupting the dorsolateral prefrontal cortex facilitates word-form learning. Scientific Reports, 7(1).

Snoeren, N. D., Gaskell, M. G., Maria, A., & Di Betta, A. M. (2009). The Perception of Assimilation in Newly Learned Novel Words. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(2), 542–549. https://doi.org/10.1037/a0014509

Spiegel, C., & Halberda, J. (2011). Journal of Experimental Child Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Child Psychology*, *109*(1), 132–140. https://doi.org/10.1016/j.jecp.2010.10.013

Stadler, M. a. (1993). Implicit serial learning: questions inspired by Hebb (1961). *Memory & Cognition*, *21*(6), 819–827. https://doi.org/10.3758/BF03202749

Staels, E., & Van den Broeck, W. (2015). No solid empirical evidence for the solid (serial order learning impairment) hypothesis of dyslexia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 650–669. https://doi.org/10.1037/xlm0000054

Swingley, D. (2010). Fast Mapping and Slow Mapping in Children's Word Learning. *Language Learning and Development*, *6*, 179–183. https://doi.org/10.1080/15475441.2010.484412

Szmalec, A., Duyck, W., Vandierendonck, A., Mata, A. B., & Page, M. P. (2009). The Hebb repetition effect as a laboratory analogue of novel word learning. *Quarterly Journal of Experimental Psychology (2006)*, *62*(July 2014), 435–443. https://doi.org/10.1080/17470210802386375

Szmalec, A., Loncke, M., Page, M. P. a, & Duyck, W. (2011). Order or disorder? Impaired

Hebb learning in dyslexia. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(5), 1270–9. https://doi.org/10.1037/a0023820

Szmalec, A., Page, M. P., & Duyck, W. (2012). The development of long-term lexical representations through Hebb repetition learning. *Journal of Memory and Language*, *67*(3), 342–354. https://doi.org/10.1016/j.jml.2012.07.001

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(43), 14356–14360. https://doi.org/10.1523/JNEUROSCI.3028-10.2010

Tham, E. K. H., Lindsay, S., & Gaskell, M. G. (2015). Markers of automaticity in sleep-associated consolidation of novel words. Neuropsychologia, 71, 146–157. https://doi.org/10.1016/j.neuropsychologia.2015.03.025

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. Cognition, 92, 231–270. https://doi.org/10.1016/j.cognition.2003.10.008

van Ooijen, B. V. A. N. (1996). Vowel mutability and lexical selection in English : *Memory*& *Cognition*, *24*(5), 573–583.

Weighall, A. R., Henderson, L. M., Barr, D. J., Cairney, S. A., & Gaskell, M. G. (2016). Eye-tracking the time-course of novel word learning and lexical competition in adults and children. *Brain and Language*, *167*, 13–27. https://doi.org/10.1016/j.bandl.2016.07.010

West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2018). The procedural learning deficit hypothesis of language learning disorders: we see some problems. Developmental science, 21(2), e12552.

Appendix A

List of Stimuli used in Experiments 1-4.

	English Base Word	Novel Word Experiments 1-3	Foils Experiments 1-3	Novel Words Experiment 4	Foils Experiment 4
List 1	celery	celedo	celemi	celero	celera
	finale	finato	finady	finalo	finaly
	recipe	recino	reciby	recipo	recipy
	bikini	bikiso	bikita	bikino	bikina
	colony	colopay	colofo	colonay	colono
	sesame	sesana	sesara	sesama	sesamy
	salary	salamo	salaky	salaro	salara
	libido	libima	libiny	libida	libidy
	cinema	cinedy	cinero	cinemy	cinemo
	casino	casira	casibu	casina	casinu
	kimono	kimota	kimore	kimona	kimone
	pagoda	pagory	pagono	pagody	pagodo
List 2	tomato	tomany	tomare	tomaty	tomate
	bakery	bakeva	bakemo	bakera	bakero
	rosary	rosano	rosava	rosaroo	rosara
	karate	karano	karaby	karato	karaty
	saliva	saliro	salika	salivo	salivu
	banana	banary	banamo	banany	banano
	safari	safano	safany	safaro	safara
	melody	meloro	melova	melodo	meloda
	sonata	sonary	sonake	sonaty	sonatay
	corona	corode	corozo	corone	corono
	canary	canato	canafy	canaro	canara
	mimosa	mimoly	mimora	mimosy	mimosay

Note. The pronunciation of the novel words and foils matched the base words on the first two syllables in terms of phonemic overlap and stress pattern. Foils 1 were used in Experiments 1-3 and oils 2 in Experiment 4.

Appendix B

Properties of Key Stimuli

List	English Base Word	Number of neighbours*	Uniqueness point	Frequency per million*	Phonological neighbours**	Orthographic neighbours**
List 1	celery	0	5	3	1	0
	finale	0	4	2	0	1
	recipe	1	4	16	0	1
	bikini	0	4	2	0	0
	colony	1	5	19	0	0
	sesame	0	4	1	0	0
	salary	0	5	29	4	0
	libido	0	4	1	0	0
	cinema	0	4	22	0	0
	casino	1	4	6	0	1
	kimono	0	3	1	0	0
	pagoda	0	3	1	0	0
List 2	tomato	0	5	15	0	0
	bakery	0	4	5	0	0
	rosary	1	4	1	1	1
	karate	0	5	5	0	0
	saliva	0	4	3	0	0
	banana	0	5	9	0	0
	safari	0	4	2	0	0
	melody	0	5	5	1	0
	sonata	0	3	3	0	0
	corona	0	5	1	0	0
	canary	1	4	3	0	0
	mimosa	0	4	1	0	0

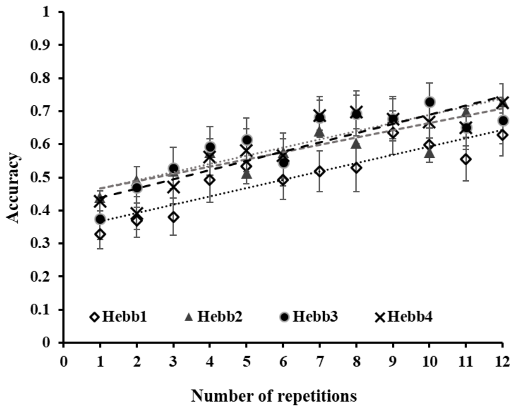
Note. The number of neighbours was checked against *CELEX using WinWordGen 1.0 (Duyck, Desmet, Verbeke, & Brysbaert, 2004) and ** English Lexicon Project (Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English Lexicon Project. Behavior Research Methods, 39, 445-459).

Supplementary Materials

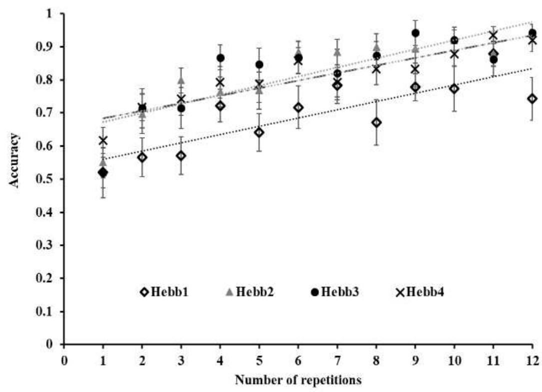
Here, we report four subsidiary analyses that provide further insights into performance across the four experiments.

1. Does accuracy vary across the four Hebb sequences in each learning phase?

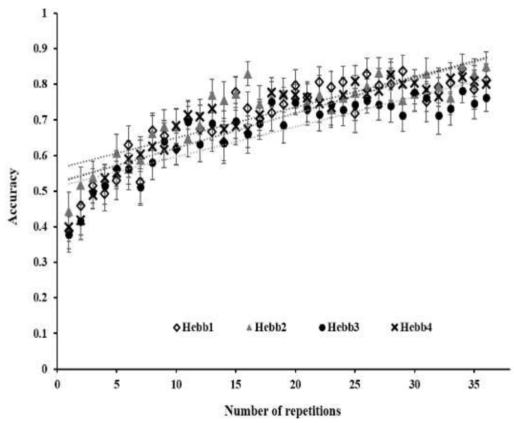
Experiment 1



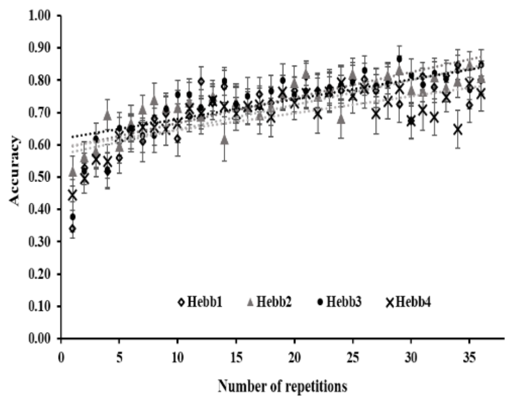
Experiment 2



Experiment 3



Experiment 4

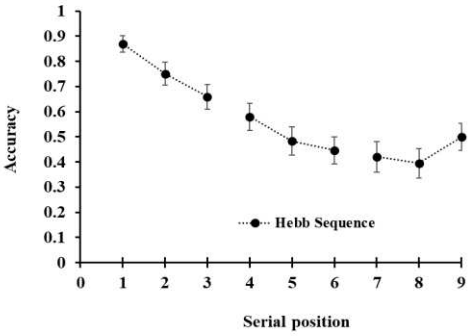


Supplementary Figure 1. Accuracy (proportion correct) across four separate Hebb sequences used in the Hebb repetition task (error bars depict standard error; regression lines illustrate the gradient of improvement in performance).

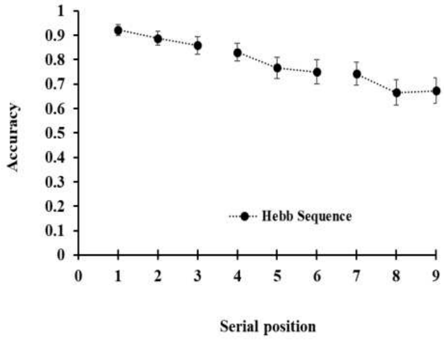
The mean number of correctly recalled syllables for each of the four Hebb lists are shown in Supplementary Figure 1. We compared the performance on different Hebb lists across our four experiments with a repeated measure analysis of variance with a within subject factor: Hebb block (1,2,3 and 4) and a mean performance on each Hebb list as our dependent variable. We found that the learning curves for four separate Hebb lists did not differ significantly in any of the four experiments (Experiment 1: $F(3, 60)=0.34$, $p=.796$, $\eta_p^2=.017$; Experiment 2: $F(3, 60)=0.29$, $p=.834$, $\eta_p^2=.014$; Experiment 3: $F(3, 87)=1.10$, $p=.356$, $\eta p^2=.036$ Experiment 4: $F(3, 87)=1.95$, $p=.128$, $\eta p^2=.063$).

2. Do temporal grouping cues alter the Hebb serial position curve?

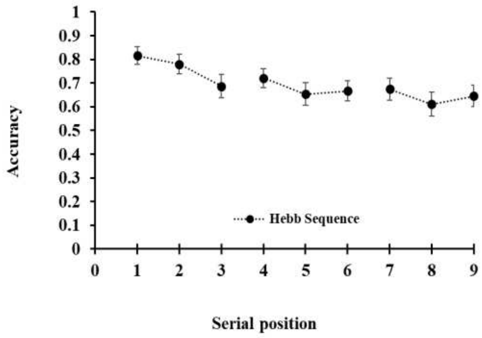
Experiment 1



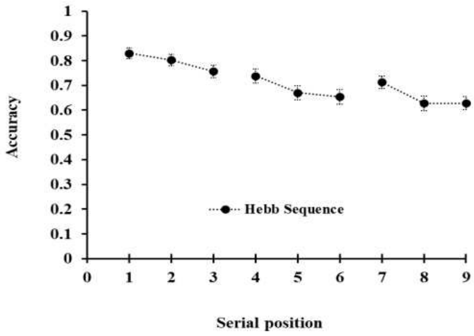
Experiment 2



Experiment 3



Experiment 4



Supplementary Figure 2. Serial position curves for Experiments 1-4. Error bars depict standard error of the means.

Accuracy in the Hebb learning task as a function of list serial position is plotted in Supplementary Figure 2. The curve for Experiment 1 (without grouping cues) is most similar to a typical Hebb serial position curve, with quite wide variations across serial position, and poorest performance in the penultimate position. The curves for Experiments 2-4 (with grouping cues) are shallower and drop off less in accuracy towards serial position 8. There is also more evidence that position within a trigram was important in these experiments. As Experiment 1 and 2 offered the best comparison of the effect of temporal grouping on immediate serial recall of Hebb sequences we subjected the recall data for these experiments to a mixed-design repeated measure ANOVA with a between-participants factor grouping (pauses versus no pauses) and a within-participants factors: trigram position (1, 2 and 3, position of the trigram within the sequence) and syllable position (position of the syllable within the trigram). The results showed that there was a reliable grouping effect ($F(1,42)=26.19$, $p<.001$, $\eta p^2=.384$), a trigram position effect ($F(2,84)=126.40$, $p<.001$, $\eta p^2=.751$) and a syllable position effect ($F(2,84)=35.18$, $p<.001$, $\eta p^2=.456$). The grouping x trigram position interaction was significant ($F(2,84)=9.36$, $p<.001$, $\eta p^2=.182$) but the grouping x syllable position interaction was not ($F(2,84)=.52$, $p=.595$, $\eta p^2=.012$). The trigram position x syllable position interaction was significant ($F(4, 168)=20.40$, $p<.001$, $\eta p^2=.327$), indicating that the recollection of a syllable depended not only on its position within a trigram but also on the position of a trigram within a sequence. The trigram position x syllable position x grouping interaction was also significant ($F(4.168)=18.76$, $p<.001$, $\eta p^2=.327$). Overall, these results indicate that the insertion of temporal pauses improved the recall of Hebb sequences and that the recall of trigrams and syllables depended on their position in the Hebb sequences. Importantly, the pause insertion differently affected the recollection of syllables depending on their position within a trigram and a sequence.

3. Does awareness of the Hebb sequences influence lexical competition?

We categorised participants in Experiments 2-4 according to whether or not they had noticed that the Hebb learning involved unannounced repeated sequences and displayed explicit knowledge of the sequence. We then used this as an independent variable in a mixed-design by-participants ANOVA with Session (0-hr, 24-hr) and Competitor Acquisition (competitor versus control), as repeated-measures variables, and Awareness (aware and not-aware) and Experiment (Experiment 2, 3 and 4) as a between-participants variable. The analysis revealed a main effect of Session ($F(1,76)=50.92$, $p<.001$, $\eta p^2=.401$ and a significant interaction between Session and Awareness ($F(1,76)=4.16$, $p=.045$, $\eta p^2=.052$. There was also a marginally significant interaction between Awareness and Experiment ($F(1,76)=3.04$, $p=.054$, $\eta p^2=.074$). Most importantly, there was no evidence that awareness influenced the Competitor Acquisition variable.

4. Do competitor effects emerge for Hebb sequences when participants successfully recall the sequences?

Across experiments, participants in the Hebb condition tended to perform more poorly in cued recall than the participants in the phoneme monitoring condition. As this difference was accompanied by a difference in the lexical integration of novel items (i.e. the phoneme monitoring group showed lexicalisation of novel items whereas the Hebb-style learning group did not) this might suggest that the relatively poor learning of novel nonwords in the Hebb condition could have affected the performance in the pause detection task. In order to address this issue we tested whether the pause detection RT data would show evidence of lexical competition when considering only the items that were correctly recalled in the cued recall task. This analysis showed no Competitor main effect or interaction with Session in any of the Experiments (see below).

Experiment 1: Session: $F_1(2,13)=12.20$, $p=.003$, $\eta p^2=.484$, $F_2(2,18)=25.72$, $p<.001$, $\eta p^2=.588$; Competitor: ($F_1(2,13)=0.42$, $p=.530$, $\eta p^2=.031$, $F_2(2,18)=.90$, $p=.355$, $\eta p^2=.05$; Session x Competitor interaction ($F_1(2,13)=.17$, $p=.844$, $\eta p^2=.013$, $F_2(2,18)=.86$, $p=.426$, $\eta p^2=.046$).

Experiment 2: Session: $F_1(2,38)=4.94$, $p=.012$, $\eta p^2=.206$, $F_2(2,38)=7.32$, $p=.002$, $\eta p^2=.278$; Competitor: ($F_1(2,38)=0.74$, $p=.399$, $\eta p^2=.038$, $F_2(2,38)=.01$, $p=.936$, $\eta p^2=.001$; Session x Competitor interaction ($F_1(2,38)=.20$, $p=.822$, $\eta p^2=.010$, $F_2(2,38)=1.53$, $p=.229$, $\eta p^2=.075$).

Experiment 3: Session: $F_1(1,28)=26.85$, $p<.001$, $\eta p^2=.490$, $F_2(1,21)=33.95$, $p=.001$, $\eta p^2=.399$; Competitor: ($F_1(1,28)=0.37$, $p=.545$, $\eta p^2=.013$, $F_2(1,21)=.31$, $p=.583$, $\eta p^2=.015$; Session x Competitor interaction ($F_1(1,28)=.16$, $p=.694$, $\eta p^2=.006$, $F_2(1,21)=.40$, $p=.533$, $\eta p^2=.019$).
Experiment 4: Session: $F_1(1,29)=2.78$, $p=.106$, $\eta p^2=.088$, $F_2(1,18)=11.95$, $p=.003$, $\eta p^2=.399$; Competitor: ($F_1(1,29)=0.36$, $p=.551$, $\eta p^2=.012$, $F_2(1,18)=.017$, $p=.688$, $\eta p^2=.009$; Session x Competitor interaction ($F_1(1,29)=.54$, $p=.471$, $\eta p^2=.018$, $F_2(1,18)=1.62$, $p=.219$, $\eta p^2=.083$).